CANADIAN
OURNAL OF MATHEMATICS

Journal Canadien de Mathématiques

VOL. III - NO. 2

1951

Published for

THE CANADIAN MATHEMATICAL CONGRESS

by the

University of Toronto Press

The chief languages of the *Journal* are English and French.

Manuscripts for publication in the *Journal* should be sent to the *Editor-in-Chief*, H. S. M. Coxeter, University of Toronto. Every paper should contain an introduction summarizing the results as far as possible in such a way as to be understood by the non-expert.

All other correspondence should be addressed to the *Managing Editor*, G. de B. Robinson, University of Toronto.

The *Journal* is published quarterly. Subscriptions should be sent to the *Managing Editor*. The price per volume of four numbers is \$6.00. This is reduced to \$3.00 for individual members of the following Societies:

Canadian Mathematical Congress
American Mathematical Society
Mathematical Association of America
London Mathematical Society
Société Mathématique de France

The Canadian Mathematical Congress gratefully acknowledges the assistance of the following towards the cost of publishing this Journal:

University of British Columbia
Ecole Polytechnique
Loyola College
McGill University
Université de Montréal
Royal Military College
Carleton College
Université Laval
University of Manitoba
McMaster University
Queen's University
University of Toronto

National Research Council of Canada
and the
American Mathematical Society
(Marion Reilly Fund)

AUTHORIZED AS SECOND CLASS MAIL, POST OFFICE DEPARTMENT, OTTAWA

ry,

in,
tz,

he
er
ole

ng

nt
is
he

he
al:

SUR LE THÉORÈME DE LEBESGUE-NIKODYM (V)

JEAN DIEUDONNÉ

Introduction. Soient E un espace compact (par exemple l'intervalle $0 \leqslant x \leqslant 1$) et μ une mesure de Radon sur E. Une fonction numérique f intégrable pour μ définit une mesure $\nu(A) = \int_A f\, d\mu$ sur l'ensemble des parties mesurables (pour μ) de E. On peut aussi considérer la forme linéaire $g \rightarrow \int gf d\mu$ qu'elle définit sur l'espace C des fonctions numériques *continues* dans E, et cette forme linéaire est continue pour la topologie définie par la norme $\|f\| = \sup_{x \in E} |f(x)|$. Si on appelle encore "mesure" une forme linéaire continue sur C, le théorème de Lebesgue-Nikodym classique caractérise celles de ces "mesures" qui sont de la forme $g \rightarrow \int gf\, d\mu$ par une condition de "continuité absolue" par rapport à μ. On définit de même une "mesure vectorielle" comme une application linéaire fortement continue $\mathbf{m}$ de C dans le dual F' d'un espace de Banach F; le problème analogue au précédent consiste à trouver les conditions moyennant lesquelles $\mathbf{m}$ peut s'écrire sous la forme $f \rightarrow \int \mathbf{g} f\, d\mu$, ou $\mathbf{g}$ est une application de E dans F', "faiblement intégrable" pour μ. Le théorème de Dunford-Pettis (voir par exemple [4]) donne une condition *suffisante* pour qu'il en soit ainsi, à savoir que l'on ait $|\mathbf{m}(f)| \leqslant a\int |f| d\mu$ pour toute fonction $f \in C$ (a constante), en supposant en outre F séparable. Nous en déduisons ici (toujours pour F séparable) la condition *nécessaire et suffisante* pour que la mesure vectorielle $\mathbf{m}$ soit de la forme voulue (th. 1).

A l'aide de ce résultat, et moyennant des hypothèses assez strictes (et peut-être superflues) de séparabilité sur E et F, nous avons pu déterminer complètement le *dual* de l'espace de Banach $L^p{}_F$ des fonctions $\mathbf{f}$ à valeurs dans F et telles que $|\mathbf{f}|^{p-1} . \mathbf{f}$ soit intégrable au sens de Bochner (th. 2); la démonstration de ce théorème semble beaucoup plus difficile pour $p > 1$ que pour $p = 1$.

1. Dans tout ce qui suit, E désigne un espace *compact*, μ une mesure de Radon positive sur E. Nous désignerons par C l'espace de Banach des fonctions réelles continues dans E (avec la norme $\|f\| = \sup_{x \in E} |f(x)|$); par L^p (pour p fini et $\geqslant 1$) l'espace des classes de fonctions réelles de puissance p-ème intégrable, avec la norme $N_p(f) = (\int |f|^p d\mu)^{1/p}$; par L^∞ l'espace des classes de fonctions mesurables et essentiellement bornées, avec la norme $N_\infty(f) = \text{ess}\, . \sup_{x \in E} |f(x)|$. Nous identifierons d'ordinaire une classe appartenant à un L^p avec une quel-

Reçu le 20 novembre, 1950.

conque des fonctions de cette classe, tout ce qui suit ne faisant intervenir des fonctions que "modulo les ensembles de mesure nulle".[1]

Dans ce qui suit, F désignera un espace de Banach, le plus souvent *séparable*, et F' son dual; pour un vecteur $\mathbf{z} \in F$ (resp. $\mathbf{z}' \in F'$), la norme de $\mathbf{z}$ (resp. $\mathbf{z}'$) sera désignée par $|\mathbf{z}|$ (resp. $|\mathbf{z}'|$). Nous allons surtout nous occuper d'applications de E dans le *dual* F' de F; pour une telle fonction $\mathbf{f}$, la notation $|\mathbf{f}|$ désignera la fonction numérique $x \to |\mathbf{f}(x)|$; pour tout vecteur fixe $\mathbf{z} \in F$, $\langle \mathbf{z}, \mathbf{f} \rangle$ sera la fonction numérique $x \to \langle \mathbf{z}, \mathbf{f}(x) \rangle$; enfin, si $\mathbf{g}$ est une application de E dans F, $\langle \mathbf{g}, \mathbf{f} \rangle$ désignera la fonction numérique $x \to \langle \mathbf{g}(x), \mathbf{f}(x) \rangle$.

On dit qu'une application $\mathbf{f}$ de E dans F' est *faiblement mesurable* si, pour tout $\mathbf{z} \in F$, la fonction numérique $\langle \mathbf{z}, \mathbf{f} \rangle$ est mesurable; nous utiliserons le lemme suivant, dû à R. Godement:

LEMME 1. *Soit F un espace de Banach séparable, et soit $\mathbf{f}$ une application faiblement mesurable de E dans F'. Alors*:

1° *la fonction numérique $|\mathbf{f}|$ est mesurable*;

2° *pour tout $\epsilon > 0$, il existe un ensemble compact $K \subset E$ tel que $\mu(E - K) \leqslant \epsilon$ et que la restriction de $\mathbf{f}$ à K soit continue pour la topologie faible $\sigma(F', F)$.*

En effet, il existe par hypothèse une suite $(\mathbf{a}_n)$ partout dense dans F. Pour tout $x \in E$, on a $|\mathbf{f}(x)| = \sup_n |\langle \mathbf{a}_n, \mathbf{f}(x) \rangle|/|\mathbf{a}_n|$; la fonction $|\mathbf{f}|$ est donc l'enveloppe supérieure d'une suite de fonctions mesurables, et par suite est mesurable. En vertu du théorème de Lusin, il existe un ensemble compact $K \subset E$ tel que $\mu(E - K) \leqslant \epsilon$ et que les restrictions à K de *toutes* les fonctions mesurables $|\mathbf{f}|$, $\langle \mathbf{a}_n, \mathbf{f} \rangle$ soient continues. Comme la restriction de $|\mathbf{f}|$ à K est bornée, pour tout $\mathbf{a} \in F$, la fonction $\langle \mathbf{a}, \mathbf{f} \rangle$ a une restriction à K qui est limite uniforme des restrictions à K d'une suite de fonctions $\langle \mathbf{a}_n, \mathbf{f} \rangle$; la restriction de $\langle \mathbf{a}, \mathbf{f} \rangle$ à K est donc continue, ce qui achève la démonstration.

Nous dirons qu'une application $\mathbf{f}$ de E dans F' est *faiblement intégrable* si, pour tout $\mathbf{z} \in F$, la fonction numérique $\langle \mathbf{z}, \mathbf{f} \rangle$ est intégrable; on sait (théorème de Gelfand-Dunford; cf. [8, p. 339]) qu'il existe alors un élément de F', appelé l'intégrale de $\mathbf{f}$ et noté $\int \mathbf{f}\, d\mu$, tel que l'on ait $\int \langle \mathbf{z}, \mathbf{f} \rangle\, d\mu = \langle \mathbf{z}, \int \mathbf{f}\, d\mu \rangle$ pour tout $\mathbf{z} \in F$; il est clair que pour toute fonction numérique $g \in L^\infty$, la fonction $\mathbf{f}g$ est encore faiblement intégrable. On a toujours $|\int \mathbf{f}\, d\mu| \leqslant \int^* |\mathbf{f}|\, d\mu$, le second membre désignant l'intégrale supérieure de $|\mathbf{f}|$, qui peut être infinie.

2. De même que, d'après le théorème classique de F. Riesz, une mesure de Radon sur E peut être définie comme forme linéaire continue sur l'espace de

[1] Au lieu d'une mesure de Radon sur un espace compact E, on pourrait naturellement considérer une mesure "abstraite" μ sur un ensemble quelconque E, telle que $\mu(E) < +\infty$. En raison du fait que les fonctions n'interviennent en réalité que par leurs classes, cette généralisation ne serait qu'apparente, puisqu'on peut alors passer à l'espace compact "de représentation" associé à la mesure μ (et dit "espace de Kakutani" dans [4]). Signalons aussi qu'on peut généraliser les résultats de ce travail au cas où E est un espace localement compact quelconque, et μ une mesure de Radon sur E.

Banach C, nous appellerons *mesure vectorielle* sur E, à valeurs dans F', toute application linéaire continue $\mathbf{m}$ de C dans F', pour la topologie forte de F', c'est-à-dire toute application linéaire satisfaisant à une inégalité de la forme

$$|\mathbf{m}(f)| \leqslant a \,.\, \|f\| \tag{1}$$

pour tout $f \in C$, a étant une constante $\geqslant 0$. Pour tout $\mathbf{z} \in F$, la forme linéaire $f \to \langle \mathbf{z}, \mathbf{m}(f)\rangle$ est continue sur C, donc de la forme $f \to \int f\, d\nu_{\mathbf{z}}$, où $\nu_{\mathbf{z}}$ est une mesure de Radon sur E.

Nous dirons que $\mathbf{m}$ est *faiblement absolument continue* par rapport à μ si chacune des mesures $\nu_{\mathbf{z}}$ est absolument continue par rapport à μ, c'est-à-dire si $\langle \mathbf{z}, \mathbf{m}(f)\rangle = \int g_{\mathbf{z}} f\, d\mu$, où $g_{\mathbf{z}}$ appartient à L^1. On peut alors prolonger $\mathbf{m}$ à L^∞ par la formule précédente: il suffit en effet de prouver que pour $f \in L^\infty$ l'application $\mathbf{z} \to \int g_{\mathbf{z}} f\, d\mu$ est une forme linéaire continue sur F pour qu'on puisse l'écrire sous la forme $\mathbf{z} \to \langle \mathbf{z}, \mathbf{m}(f)\rangle$, où $\mathbf{m}(f)$ est un élément bien déterminé de F'. Or, pour $f \in C$, on a, d'après (1), l'inégalité

$$|\langle \mathbf{z}, \mathbf{m}(f)\rangle| = \left|\int g_{\mathbf{z}} f\, d\mu\right| \leqslant a \,.\, \|f\| \,.\, |\mathbf{z}| \tag{2}$$

d'où résulte

$$\int |g_{\mathbf{z}}|\, d\mu = \sup_{f \in C} \left|\int g_{\mathbf{z}} f\, d\mu\right| / \|f\| \leqslant a \,.\, |\mathbf{z}|,$$

et par suite on a encore $\left|\int g_{\mathbf{z}} f\, d\mu\right| \leqslant a \,.\, N_\infty(f) \,.\, |\mathbf{z}|$ pour toute fonction $f \in L^\infty$, ce qui établit notre assertion et montre en même temps que le prolongement de $\mathbf{m}$ à L^∞ satisfait à l'inégalité

$$|\mathbf{m}(f)| \leqslant a \,.\, N_\infty(f). \tag{3}$$

Nous supposerons toujours désormais que les mesures vectorielles faiblement absolument continues que nous considérerons sont prolongées de cette manière à L^∞. L'exemple le plus important de telles mesures est fourni par les applications de la forme $f \to \int \mathbf{g} f\, d\mu$, où $\mathbf{g}$ est une application *faiblement intégrable* de E dans F'; on sait en effet alors que l'application $\mathbf{z} \to \langle \mathbf{z}, \mathbf{g}\rangle$ de F dans L^1 est continue [8, p. 339], autrement dit qu'on a

$$\int |\langle \mathbf{z}, \mathbf{g}\rangle|\, d\mu \leqslant b \,.\, |\mathbf{z}| \tag{4}$$

(b constante $\geqslant 0$); on en déduit que pour toute fonction $f \in L^\infty$, on a

$$\left|\int \langle \mathbf{z}, \mathbf{g} f\rangle d\mu\right| \leqslant N_\infty(f) \,.\, \int |\langle \mathbf{z}, \mathbf{g}\rangle| d\mu \leqslant b \,.\, |\mathbf{z}| \,.\, N_\infty(f),$$

et par suite

$$\left|\int \mathbf{g} f\, d\mu\right| \leqslant b \,.\, N_\infty(f). \tag{5}$$

Le problème se pose de *caractériser* les mesures vectorielles qui peuvent se mettre sous la forme $\mathbf{m}(f) = \int \mathbf{g} f\, d\mu$: c'est une généralisation du problème analogue pour les mesures réelles, dont la solution est donnée par le théorème de Lebesgue-Nikodym. S'inspirant de ce dernier résultat, on peut se demander si la condition cherchée ne serait pas obtenue en exprimant que $\mathbf{m}$ est *forte-*

ment absolument continue: on entend par là que, pour tout $\epsilon > 0$, il existe $\delta > 0$ tel que les relations $f \in L^\infty$, $|f| \leqslant 1$, $N_1(f) \leqslant \delta$ entraînent $|\mathbf{m}(f)| \leqslant \epsilon$. Mais on connaît des exemples de mesures vectorielles fortement absolument continues et qui ne sont pas de la forme $\int \mathfrak{g} f \, d\mu$ [9, p. 303, exemple 9.4], et aussi des exemples de mesures vectorielles de la forme $\int \mathfrak{g} f \, d\mu$ qui ne sont pas fortement absolument continues [2, p. 377, exemple 7]. La solution doit donc être cherchée dans une autre direction.

Etant donné un ensemble compact $K \subset E$, nous dirons que la mesure vectorielle $\mathbf{m}$ est *majorée par un multiple de* μ *dans* K s'il existe une constante $a_K \geqslant 0$ telle que, pour toute fonction $f \in L^\infty$, nulle dans $E - K$, on a $|\mathbf{m}(f)| \leqslant a_K \, . \, N_1(f)$.

THÉORÈME 1. *L'espace F étant séparable, pour qu'une mesure vectorielle $\mathbf{m}$, à valeurs dans F', faiblement absolument continue par rapport à μ, soit de la forme $f \to \int \mathfrak{g} f \, d\mu$, où $\mathfrak{g}$ est faiblement intégrable, il faut et il suffit que, pour tout $\epsilon > 0$, il existe un ensemble compact $K \subset E$ tel que $\mu(E - K) \leqslant \epsilon$ et que, dans K, $\mathbf{m}$ soit majorée par un multiple de μ; en outre, la fonction $\mathfrak{g}$ telle que $\mathbf{m}(f) = \int \mathfrak{g} f \, d\mu$, est déterminée à un ensemble de mesure nulle près.*

La condition est *nécessaire*: en effet, si $\mathbf{m}(f) = \int \mathfrak{g} f \, d\mu$, $\mathfrak{g}$ est faiblement mesurable, donc il résulte du lemme 1 que $|\mathfrak{g}|$ est mesurable et à valeurs finies dans E. Pour tout $\epsilon > 0$, il existe donc un ensemble compact $K \subset E$ tel que $\mu(E - K) \leqslant \epsilon$, et que $|\mathfrak{g}|$ soit bornée dans K. Mais si $|\mathfrak{g}(x)| \leqslant a_K$ pour $x \in K$, il est clair que pour toute fonction $f \in L^\infty$, nulle dans $E - K$, on a

$$|\int \mathfrak{g} f \, d\mu| = |\int_K \mathfrak{g} f \, d\mu| \leqslant a_K \, . \int_K |f| \, d\mu = a_K \, . \, N_1(f).$$

La condition est *suffisante*. On peut en effet, par hypothèse, trouver une suite croissante d'ensembles compacts $K_n \subset E$ tels que $\mu(E - K_n)$ tende vers 0 et que, dans chaque K_n, $\mathbf{m}$ soit majorée par un multiple de μ. Soit V_n le sous-espace de L^∞ formé des fonctions nulles hors de K_n; il existe donc une constante $a_n \geqslant 0$ telle que, pour toute fonction $f \in V_n$, on ait $|\mathbf{m}(f)| \leqslant a_n \, . \, N_1(f)$. Il résulte alors du théorème de Dunford-Pettis (voir par exemple [4] ou [5]) qu'il existe une fonction $\mathfrak{g}_n$ à valeurs dans F', faiblement intégrable dans K_n et telle que $\mathbf{m}(f) = \int \mathfrak{g}_n f \, d\mu$ pour toute fonction $f \in V_n$. Montrons que $\mathfrak{g}_n$ est égale presque partout à la restriction de toute fonction $\mathfrak{g}_m$ à K_n pour $m \geqslant n$. En effet, si $(\mathbf{a}_r)$ est une suite d'éléments de F, partout dense dans F, on a pour toute fonction $f \in V_n$ et tout indice r,

$$\langle \mathbf{a}_r, \mathbf{m}(f) \rangle = \int \langle \mathbf{a}_r, \mathfrak{g}_n \rangle f \, d\mu = \int \langle \mathbf{a}_r, \mathfrak{g}_m \rangle f \, d\mu,$$

ce qui prouve que, dans K_n, $\langle \mathbf{a}_r, \mathfrak{g}_n(x) \rangle = \langle \mathbf{a}_r, \mathfrak{g}_m(x) \rangle$ presque partout; on en déduit que cette relation a aussi lieu *pour tout* r presque partout dans K_n, ce qui prouve que $\mathfrak{g}_m(x) = \mathfrak{g}_n(x)$ presque partout dans K_n.

Il existe donc une fonction $\mathfrak{g}$ à valeurs dans F', définie dans E, et dont la restriction à K_n est presque partout égale à $\mathfrak{g}_n$, ce qui prouve que, pour tout $f \in V_n$, $\mathfrak{g} f$ est faiblement intégrable et $\mathbf{m}(f) = \int \mathfrak{g} f \, d\mu$. Nous allons en déduire

que $\mathbf{g}$ est faiblement intégrable (dans E tout entier) et qu'on a $\mathbf{m}(f) = \int \mathbf{g} f\, d\mu$ pour toute fonction $f \in L^\infty$. En effet, soit $\mathbf{z}$ quelconque dans F; pour tout entier n, on a

$$\langle \mathbf{z}, \mathbf{m}\, (f\phi_{K_n})\rangle = \int \langle \mathbf{z}, \mathbf{g}\phi_{K_n}\rangle f\, d\mu.$$

Comme par hypothèse $\mathbf{m}$ est faiblement absolument continue, $\langle \mathbf{z}, \mathbf{m}(f\phi_{K_n})\rangle$ tend vers $\langle \mathbf{z}, \mathbf{m}(f)\rangle$ lorsque n croît indéfiniment. Cela signifie que la suite de fonctions intégrables $\langle \mathbf{z}, \mathbf{g}_n\rangle = \langle \mathbf{z}, \mathbf{g}\phi_{K_n}\rangle$ est une *suite de Cauchy* pour la topologie faible dans L^1; on sait qu'une telle suite est faiblement convergente, et comme en outre $\langle \mathbf{z}, \mathbf{g}_n(x)\rangle$ tend *partout* vers $\langle \mathbf{z}, \mathbf{g}(x)\rangle$, la suite des $\langle \mathbf{z}, \mathbf{g}_n\rangle$ est *fortement convergente* vers $\langle \mathbf{z}, \mathbf{g}\rangle$ dans L^1 (voir par exemple [6]). En passant à la limite, on voit que $\langle \mathbf{z}, \mathbf{g}\rangle$ est intégrable et que pour tout $f \in L^\infty$,

$$\langle \mathbf{z}, \mathbf{m}(f)\rangle = \int \langle \mathbf{z}, \mathbf{g} f\rangle d\mu.$$

L'unicité de $\mathbf{g}$ (à un ensemble de mesure nulle près) se démontre comme ci-dessus le fait que $\mathbf{g}_m$ et $\mathbf{g}_n$ sont égales presque partout dans K_n pour $m \geqslant n$.

3. Rappelons qu'une application $\mathbf{f}$ de E dans un espace de Banach G est dite *fortement intégrable* si $|\mathbf{f}|$ est intégrable, et s'il existe une suite $(\mathbf{f}_n)$ de fonctions étagées (fonctions mesurables ne prenant qu'un nombre fini de valeurs) telle que $\int |\mathbf{f} - \mathbf{f}_n| d\mu$ tende vers 0. On pose alors $N_1(\mathbf{f}) = \int |\mathbf{f}| d\mu$, et l'espace vectoriel L^1_G des fonctions fortement intégrables (identifiées lorsqu'elles ne diffèrent que sur un ensemble de mesure nulle), muni de la norme $N_1(\mathbf{f})$, est un *espace de Banach*. Pour tout $p > 1$, on définit de même l'espace L^p_G comme l'espace des applications $\mathbf{f}$ de E dans G telles que $|\mathbf{f}|^{p-1}.\mathbf{f}$ soit fortement intégrable; muni de la norme $N_p(\mathbf{f}) = (\int |\mathbf{f}|^p d\mu)^{1/p}$, c'est encore un espace de Banach, dans lequel le sous-espace P_G des fonctions étagées est partout dense. On déduit aussitôt de là que si l'espace L^1 (des fonctions intégrables *numériques*) et l'espace G sont tous deux séparables, l'espace L^p_G $(1 \leqslant p < +\infty)$ est aussi *séparable*.

Dans ce qui suit, p et q désignent deux exposants conjugués $\geqslant 1$ c'est-à-dire tels que $p^{-1} + q^{-1} = 1$. Si $\mathbf{f}$ est une fonction de L^p_F, $\mathbf{g}$ une fonction de $L^q_{F'}$, $\langle \mathbf{f}, \mathbf{g}\rangle$ est une fonction numérique intégrable et on a l'inégalité de Hölder

$$\int |\langle \mathbf{f}, \mathbf{g}\rangle| d\mu \leqslant N_p(\mathbf{f}) N_q(\mathbf{g}). \tag{6}$$

Cette inégalité peut en outre être précisée par le lemme suivant:

LEMME 2. *Pour* $1 < p < +\infty$, *on a les relations*

$$N_p(\mathbf{f}) = \sup_{N_q(\mathbf{g}) \leqslant 1} \int \langle \mathbf{f}, \mathbf{g}\rangle d\mu \tag{7}$$

$$N_q(\mathbf{g}) = \sup_{N_p(\mathbf{f}) \leqslant 1} \int \langle \mathbf{f}, \mathbf{g}\rangle d\mu. \tag{8}$$

Démontrons par exemple la relation (8). Prenons d'abord le cas particulier où $\mathbf{g}$ est une fonction étagée, $\mathbf{g} = \sum_k \mathbf{a}'_k \phi_{A_k}$, où les A_k forment une partition

finie de E en ensembles mesurables; on peut supposer en outre, en multipliant $\mathfrak{g}$ par un scalaire, que $N_q(\mathfrak{g}) = 1$, c'est-à-dire $\sum_k |\mathbf{a}'_k|^q . \mu(A_k) = 1$. Pour une fonction étagée $\mathbf{f}$ de la forme $\mathbf{f} = \sum_k \mathbf{a}_k \phi_{A_k}$ telle que $N_p(\mathbf{f}) \leqslant 1$, on a $\sum_k |\mathbf{a}_k|^p . \mu(A_k) \leqslant 1$ et

$$\int \langle \mathbf{f}, \mathfrak{g} \rangle d\mu = \sum_k \langle \mathbf{a}_k, \mathbf{a}'_k \rangle . \mu(A_k).$$

Or, par définition de la norme dans F', on peut, pour chaque k, trouver $\mathbf{a}_k$ tel que $|\mathbf{a}_k|^p = |\mathbf{a}'_k|^q$ et que $\langle \mathbf{a}_k, \mathbf{a}'_k \rangle$ soit arbitrairement voisin de $|\mathbf{a}'_k|^{1+q/p} = |\mathbf{a}'_k|^q$; la relation (8) est donc vraie dans ce cas.

Passons au cas général, et supposons encore que $N_q(\mathfrak{g}) = 1$. Il existe alors une fonction étagée $\mathfrak{g}_1$ telle que $N_q(\mathfrak{g} - \mathfrak{g}_1) \leqslant \epsilon$, d'où en particulier $N_q(\mathfrak{g}_1) \geqslant 1 - \epsilon$; d'après la première partie de la démonstration, il existe une fonction $\mathbf{f} \in L^p_F$ telle que $N_p(\mathbf{f}) \leqslant 1$ et

$$\int \langle \mathbf{f}, \mathfrak{g}_1 \rangle d\mu \geqslant N_q(\mathfrak{g}_1) - \epsilon \geqslant 1 - 2\epsilon;$$

mais d'autre part, on a, d'après (6)

$$\left| \int \langle \mathbf{f}, \mathfrak{g} - \mathfrak{g}_1 \rangle d\mu \right| \leqslant N_p(\mathbf{f}) N_q(\mathfrak{g} - \mathfrak{g}_1) \leqslant \epsilon,$$

d'où $\int \langle \mathbf{f}, \mathfrak{g} \rangle d\mu \geqslant 1 - 3\epsilon$, ce qui achève de démontrer le lemme (le raisonnement étant tout à fait analogue pour (7)).

Les relations (6) et (8) prouvent que l'application $\mathbf{f} \to \int \langle \mathbf{f}, \mathfrak{g} \rangle d\mu$ est une *forme linéaire continue* sur l'espace de Banach L^p_F, et que la norme de cette forme linéaire est égale à $N_q(\mathfrak{g})$. On peut donc identifier $L^q_{F'}$ à un sous-espace (fortement fermé) du dual $(L^p_F)'$ de L^p_F.

4. Nous allons maintenant déterminer complètement le dual de L^p_F moyennant des hypothèses supplémentaires de séparabilité.

THÉORÈME 2. *Les espaces de Banach F et L^1 étant supposés séparables, toute forme linéaire continue sur L^p_F $(1 < p < +\infty)$ peut s'écrire $\mathbf{f} \to \int \langle \mathbf{f}, \mathfrak{g} \rangle d\mu$, où $\mathfrak{g}$ est faiblement mesurable et telle que $N_q(|\mathfrak{g}|) < +\infty$; $\mathfrak{g}$ est en outre déterminée à un ensemble de mesure nulle près. Inversement, pour toute fonction $\mathfrak{g}$ ayant ces propriétés, $\mathbf{f} \to \int \langle \mathbf{f}, \mathfrak{g} \rangle d\mu$ est une forme linéaire continue sur L^p_F, dont la norme est égale à $N_q(|\mathfrak{g}|)$.*

J'ai démontré un théorème analogue pour le cas $p = 1$, en supposant seulement F séparable [4 et 5]; nous allons suivre la même méthode, mais son application sera beaucoup plus délicate que pour $p = 1$.

Soit θ une forme linéaire continue sur L^p_F, et soit a sa norme; on a donc

$$|\theta(\mathbf{f})| \leqslant a . N_p(\mathbf{f}) \tag{9}$$

pour tout $\mathbf{f} \in L^p_F$. Pour tout vecteur $\mathbf{z} \in F$, et toute fonction numérique $f \in L^p$, $\mathbf{z} \to \theta(f\mathbf{z})$ est une forme linéaire sur F, telle que

$$|\theta(f\mathbf{z})| \leqslant a . |\mathbf{z}| . N_p(f). \tag{10}$$

Autrement dit, cette forme linéaire est continue, et on peut écrire $\theta(f\mathbf{z}) = \langle \mathbf{z}, \mathbf{m}(f) \rangle$, où $\mathbf{m}(f) \in F'$ et est telle que

$$|\mathbf{m}(f)| \leqslant a \,.\, N_p(f) \tag{11}$$

pour toute fonction $f \in L^p$. Il résulte de là que $\mathbf{m}$, restreinte à L^∞, est une mesure vectorielle à valeurs dans F', faiblement absolument continue par rapport à μ. Mais tandis que, pour $p = 1$, l'inégalité (11) permet aussitôt d'appliquer à $\mathbf{m}$ le théorème de Dunford-Pettis, et de montrer que $\mathbf{m}(f)$ est de la forme $\int \mathbf{g} f \, d\mu$, il n'en est plus de même pour $p > 1$; on connaît en effet des exemples de mesures vectorielles qui satisfont à une inégalité telle que (11) et qui ne sont pas de la forme $\int \mathbf{g} f \, d\mu$, lorsque $p > 1$ [9, p. 303, exemple 9.4]. Nous allons chercher à appliquer à $\mathbf{m}$ le théorème 1, en précisant la nature de la forme linéaire θ.

Pour cela, remarquons que, d'après ce qui a été vu au n° 3, $L^q{}_{F'}$ est un sous-espace de $(L^p{}_F)'$; en outre, pour toute fonction $\mathbf{f} \in L^p{}_F$, il résulte du lemme 2 que, pour tout $\epsilon > 0$, il existe une fonction $\mathbf{g} \in L^q{}_{F'}$ telle que $N_q(\mathbf{g}) = 1$ et $\int \langle \mathbf{f}, \mathbf{g} \rangle d\mu \geqslant (1 - \epsilon) N_p(\mathbf{f})$. Comme $L^p{}_F$ est *séparable* en vertu des hypothèses, un théorème de Banach, précisé par J. Dixmier [7, th. 1 et 7] montre qu'il existe une suite $(\mathbf{g}_n)$ de fonctions de $L^q{}_{F'}$, telles que $N_q(\mathbf{g}_n) \leqslant a$ pour tout n, et que la suite des formes linéaires $\mathbf{f} \to \int \langle \mathbf{f}, \mathbf{g}_n \rangle d\mu$ tende *faiblement* vers θ, c'est-à-dire que pour toute fonction $\mathbf{f} \in L^p{}_F$, $\int \langle \mathbf{f}, \mathbf{g}_n \rangle d\mu$ tend vers $\theta(\mathbf{f})$.

Nous allons chercher à modifier la suite $(\mathbf{g}_n)$ de façon à "améliorer" cette convergence, tout en gardant les propriétés précédentes. Considérons la suite des fonctions numériques $|\mathbf{g}_n|$; elles appartiennent à L^q, et on a $N_q(|\mathbf{g}_n|) \leqslant a$. Comme l'espace L^q est réflexif, la boule $N_q(g) \leqslant a$ est faiblement compacte dans L^q; en outre, L^p étant séparable, la topologie faible sur cette boule est métrisable, donc on peut extraire de la suite $(|\mathbf{g}_n|)$ une suite qui est *faiblement convergente* vers une fonction h; nous supposerons cette extraction déjà faite. Un théorème de Banach et Mazur [1, p. 246] montre alors qu'il existe une suite de fonctions

$$h_n = \sum_{k=0}^{\infty} c_{nk} |\mathbf{g}_{n+k}|$$

où les c_{nk} sont des nombres $\geqslant 0$, nuls sauf pour un nombre fini d'indices k (dépendant de n), et tels que $\sum_k c_{nk} = 1$, qu'on peut choisir de sorte que la suite (h_n) *converge fortement*[2] dans L^q vers h. Nous allons alors remplacer $\mathbf{g}_n$ par la fonction

$$\mathbf{g}'_n = \sum_{k=0}^{\infty} c_{nk} \mathbf{g}_{n+k}$$

[2] L'énoncé que nous utilisons ici étant un peu plus précis que celui que cite Banach (*loc. cit.*), rappelons-en rapidement la démonstration. Comme h est faiblement adhérente à l'enveloppe convexe des $|\mathbf{g}_{n+k}|$ $(k \geqslant 0)$, elle est aussi fortement adhérente à cette enveloppe convexe; autrement dit, on peut déterminer les c_{nk} de sorte que $N_q(h - h_n) \leqslant 1/n$, d'où la proposition.

pour tout n; comme $\int\langle \mathbf{f}, \mathbf{g}'_n\rangle d\mu = \sum_{k=0}^{\infty} c_{nk} \int\langle \mathbf{f}, \mathbf{g}_{n+k}\rangle d\mu$, la suite $(\mathbf{g}'_n)$ est encore faiblement convergente vers θ dans $(L^p{}_F)'$; en outre on a $|\mathbf{g}'_n| \leqslant h_n$, d'où $N_q(\mathbf{g}'_n) \leqslant N_q(h_n) \leqslant \sum_{k=0}^{\infty} c_{nk} N_q(\mathbf{g}_{n+k}) \leqslant a$.

Cela étant, il existe une suite extraite de (h_n) et qui converge *presque partout* vers h dans E; nous supposerons encore qu'on ait fait cette extraction. En vertu du théorème d'Egoroff, pour tout $\epsilon > 0$, il existe un ensemble compact $K \subset E$ tel que $\mu(E - K) \leqslant \epsilon$ et que la suite (h_n) converge *uniformément* vers h dans K; on peut supposer en outre que la restriction à K des h_n et de h est continue (théorème de Lusin), d'où résulte que les fonctions h_n sont *uniformément bornées* dans K. En d'autres termes, il existe une constante b_K telle que $h_n(x) \leqslant b_K$ dans K, et par suite $|\mathbf{g}'_n(x)| \leqslant b_K$ dans K; pour tout $\mathbf{z} \in F$ et toute $f \in L^\infty$ nulle dans $E - K$, on a donc

$$|\langle \mathbf{z}, \int \mathbf{g}'_n f\, d\mu\rangle| \leqslant b_K \,.\, |\mathbf{z}| \,.\, N_1(f)$$

d'où, en passant à la limite

$$|\langle \mathbf{z}, \mathbf{m}(f)\rangle| \leqslant b_K \,.\, |\mathbf{z}| \,.\, N_1(f)$$

ce qui donne $|\mathbf{m}(f)| \leqslant b_K \,.\, N_1(f)$. En d'autres termes, nous avons prouvé que la mesure vectorielle $\mathbf{m}$ satisfait aux hypothèses du th. 1, d'où l'existence et l'unicité (à un ensemble de mesure nulle près) d'une fonction faiblement intégrable $\mathbf{g}$ à valeurs dans F', telle que $\mathbf{m}(f) = \int \mathbf{g} f\, d\mu$.

Il n'est pas évident que les fonctions $\mathbf{g}'_n$ tendent presque partout faiblement (c'est-à-dire pour la topologie faible $\sigma(F', F)$) vers $\mathbf{g}$; mais nous allons voir qu'on peut aussi réaliser cette "amélioration" par une nouvelle modification de la suite $(\mathbf{g}'_n)$. Soit $(\mathbf{a}_m)$ une suite partout dense dans la boule $|\mathbf{z}| \leqslant 1$ de F. Pour tout m et toute fonction $f \in L^p$, $\int\langle \mathbf{a}_m, \mathbf{g}'_n\rangle f\, d\mu$ tend par hypothèse vers $\langle \mathbf{a}_m, \mathbf{m}(f)\rangle$; comme dans L^q, toute suite de Cauchy pour la topologie faible est faiblement convergente, cela signifie que dans L^q, la suite des fonctions $\langle \mathbf{a}_m, \mathbf{g}'_n\rangle$ est faiblement convergente. En outre, sa limite n'est autre que $\langle \mathbf{a}_m, \mathbf{g}\rangle$ car pour toute fonction $f \in L^\infty$, $\int\langle \mathbf{a}_m, \mathbf{g}'_n\rangle f\, d\mu$ tend vers $\int\langle \mathbf{a}_m, \mathbf{g}\rangle f\, d\mu$, et L^∞ est dense dans L^p pour la topologie forte de L^p.

Cela étant, nous allons définir une suite double $\mathbf{g}_n^{(m)}$ de fonctions de $L^q{}_{F'}$, par la récurrence suivante:

$$\mathbf{g}_n^{(0)} = \mathbf{g}'_n \text{ et } \mathbf{g}_n^{(m)} = \sum_{k=0}^{\infty} c_{nk}^{(m)} \mathbf{g}_{n+k}^{(m-1)}$$

ou les $c_{nk}^{(m)}$ sont des nombres $\geqslant 0$, nuls sauf pour un nombre fini d'indices k (dépendant de m et n), et tels que $\sum_{k=0}^{\infty} c_{nk}^{(m)} = 1$, choisis de sorte que la suite des fonctions $\langle \mathbf{a}_m, \mathbf{g}_n^{(m)}\rangle$ converge *fortement* dans L^q vers $\langle \mathbf{a}_m, \mathbf{g}\rangle$. Pour prouver que ce choix est possible, il suffit, en vertu du théorème de Banach-

Mazur rappelé plus haut, de prouver que la suite $\langle \mathbf{a}_m, \mathbf{g}_n^{(m-1)} \rangle$ converge *faiblement* dans L^q vers $\langle \mathbf{a}_m, \mathbf{g} \rangle$. Or, il est immédiat, par récurrence sur m, que $\mathbf{g}_n^{(m-1)}$ appartient à l'enveloppe convexe des $\mathbf{g}'_{n+k}$ ($k \geqslant 0$), d'où la propriété cherchée.

Considérons maintenant la suite "diagonale" $\mathbf{g}''_n = \mathbf{g}_n^{(n)}$; il est clair que, pour tout indice m, $\mathbf{g}''_n$ appartient à l'enveloppe convexe des $\mathbf{g}_{n+k}^{(m)}$ ($k \geqslant 0$) pour tout $n \geqslant m$; la suite $\langle \mathbf{a}_m, \mathbf{g}''_n \rangle$ converge donc *fortement* vers $\langle \mathbf{a}_m, \mathbf{g} \rangle$ dans L^q pour *tout* indice m. Par extraction répétée de suites et utilisation du procédé diagonal, on peut supposer en outre qu'il existe un ensemble de mesure nulle H tel que pour tout $x \in E-H$ et *tout* m, la suite $\langle \mathbf{a}_m, \mathbf{g}''_n(x) \rangle$ converge vers $\langle \mathbf{a}_m, \mathbf{g}(x) \rangle$. Mais $|\mathbf{g}''_n(x)|$ est borné par un nombre de l'enveloppe convexe des nombres $|\mathbf{g}'_{n+k}(x)|$ ($k \geqslant 0$); d'après ce qu'on a vu ci-dessus, on peut toujours supposer (en agrandissant au besoin H) que l'ensemble H est tel que pour tout $x \in E - H$ la suite des $|\mathbf{g}''_n(x)|$ soit *bornée*. Comme les $\mathbf{a}_m$ forment un ensemble dense dans la boule $|\mathbf{z}| \leqslant 1$ de F, on conclut de là que, pour tout $x \in E - H$, la suite $(\mathbf{g}''_n(x))$ tend faiblement vers $\mathbf{g}(x)$. On a par suite presque partout

$$|\mathbf{g}(x)| \leqslant \liminf_{n \to \infty} |\mathbf{g}''_n(x)|;$$

mais on vient de voir que

$$|\mathbf{g}''_n(x)| \leqslant \sup_{k \geq 0} |\mathbf{g}'_{n+k}(x)| \leqslant \sup_{k \geq 0} h_{n+k}(x),$$

et par suite

$$\liminf_{n \to \infty} |\mathbf{g}''_n(x)| \leqslant \limsup_{n \to \infty} h_n(x) = h(x).$$

On a ainsi prouvé que $|\mathbf{g}(x)| \leqslant h(x)$ presque partout, d'où $N_q(|\mathbf{g}|) \leqslant N_q(h) \leqslant a$.

Donnons-nous inversement une fonction $\mathbf{g}$ à valeurs dans F', faiblement intégrable et telle que $N_q(|\mathbf{g}|) < +\infty$. Pour toute fonction $\mathbf{f} \in L^p_F$, $\langle \mathbf{f}, \mathbf{g} \rangle$ est une fonction numérique mesurable: en effet il résulte du lemme 1 et du théorème de Lusin que, pour tout $\epsilon > 0$, il existe un ensemble compact $K \subset E$ tel que $\mu(E - K) \leqslant \epsilon$ et que la restriction de $\mathbf{f}$ à K soit fortement continue, et la restriction de $\mathbf{g}$ à K faiblement continue; on en déduit aussitôt que la restriction de $\langle \mathbf{f}, \mathbf{g} \rangle$ à K est continue, donc que $\langle \mathbf{f}, \mathbf{g} \rangle$ est mesurable.[3] En outre, on a $|\langle \mathbf{f}, \mathbf{g} \rangle| \leqslant |\mathbf{f}| \,.\, |\mathbf{g}|$, d'où, par l'inégalité de Hölder,

$$\int^* |\langle \mathbf{f}, \mathbf{g} \rangle| \, d\mu \leqslant N_p(\mathbf{f}) N_q(|\mathbf{g}|),$$

ce qui prouve que $\langle \mathbf{f}, \mathbf{g} \rangle$ est intégrable et que l'application $\mathbf{f} \to \int \langle \mathbf{f}, \mathbf{g} \rangle d\mu$ de L^p_F dans $\mathbf{R}$ est une forme linéaire continue de norme $\leqslant N_q(|\mathbf{g}|)$. Si nous revenons alors à la détermination de la forme linéaire θ, pour toute fonction $\mathbf{f} \in L^p_F$ de

[3] Ce résultat et sa démonstration sont dûs à R. Godement.

la forme $\mathbf{f} = \sum_k \mathbf{c}_k f_k$, ou les $\mathbf{c}_k \in F$ et les $f_k \in L^p$, on a

$$\theta(\mathbf{f}) = \sum_k \theta(\mathbf{c}_k f_k) = \sum_k \langle \mathbf{c}_k, \mathbf{m}(f_k)\rangle = \sum_k \langle \mathbf{c}_k, \int \mathbf{g} f_k d\mu\rangle$$
$$= \int \langle \sum_k \mathbf{c}_k f_k, \mathbf{g}\rangle d\mu = \int \langle \mathbf{f}, \mathbf{g}\rangle\, d\mu.$$

Or, les combinaisons linéaires de fonctions de L^p à coefficients dans F forment un ensemble partout dense dans $L^p{}_F$; comme les deux membres extrêmes des égalités précédentes sont fonctions continues dans $L^p{}_F$ et sont identiques dans cet ensemble partout dense, on a $\theta(\mathbf{f}) = \int \langle \mathbf{f}, \mathbf{g}\rangle d\mu$ pour toute fonction $\mathbf{f} \in L^p{}_F$; en outre, on a $a \leqslant N_q(|\mathbf{g}|)$; mais comme nous avons démontré l'inégalité opposée, on a bien $a = N_q(|\mathbf{g}|)$ et le théorème 2 est complètement démontré.

5. Avec les mêmes hypothèses de séparabilité que dans le th. 2, supposons en outre que l'espace F soit *réflexif*. Alors F', dont le dual F est séparable, est lui-même séparable [1, p. 189]; il résulte d'un théorème de Pettis [9, p. 278] que la fonction $\mathbf{g}$ du th. 2 est (fortement) mesurable, et par suite que $\mathbf{g} \in L^q{}_{F'}$; en d'autres termes, l'espace $L^p{}_F$ est alors *réflexif* et a pour dual $L^q{}_{F'}$; ce résultat était connu dans le cas où F est supposé *uniformément convexe*, et même alors sans hypothèse de séparabilité [3] sur F ou L^1.

Lorsque F n'est pas réflexif, en général le dual de $L^p{}_F$ n'est pas identique à $L^q{}_{F'}$: il suffit en effet de considérer les cas où il existe une fonction $\mathbf{g}$ à valeurs dans F' faiblement mesurable, essentiellement bornée et non fortement mesurable; alors la forme linéaire $\mathbf{f} \to \int \langle \mathbf{f}, \mathbf{g}\rangle d\mu$ sera continue dans chacun des $L^p{}_F$, mais n'appartiendra à aucun des $L^q{}_{F'}$. On a un tel exemple en prenant pour E l'intervalle $0 \leqslant x \leqslant 1$, pour μ la mesure de Lebesgue sur E, pour F l'espace $(l^{(1)})$ de Banach (espaces des séries absolument convergentes), dont le dual est l'espace (m) des suites bornées; à tout $x \in E$ on fait alors correspondre la suite $\mathbf{g}(x)$ des chiffres du *développement dyadique propre* de x. On a évidemment $|\mathbf{g}(x)| \leqslant 1$ et on vérifie facilement que $\mathbf{g}$ est faiblement mesurable; mais $\mathbf{g}$ n'est pas fortement mesurable, car la relation $x \neq y$ entraîne $|\mathbf{g}(x) - \mathbf{g}(y)| = 1$ et par suite l'image par $\mathbf{g}$ d'une partie non dénombrable quelconque de E n'est jamais contenue dans un sous-espace séparable de F'.

Je signale à cette occasion un lapsus dans mon article *Sur le théorème de Lebesgue-Nikodym* (II) (Bull. Soc. Math. de France, t. 72 (1944), p. 193-239); la raison invoquée dans les lignes 9 et 10 à partir du bas de la p. 219 n'est bien entendu pas correcte. Il faut dire que l'application $u \to u^2$ est continue dans toute partie de $L(\Omega)$ formée des classes de fonctions intégrables u telles que $|u| \leqslant v$, où v est une fonction intégrable ainsi que v^2; en effet, $|\int (u^2 - u_0^2)d\mu| \leqslant 2\int v|u - u_0|d\mu$; or, pour tout $\epsilon > 0$, il existe un entier n tel que dans l'ensemble A des points où $v(x) \geqslant n$, on ait $\int_A v^2 d\mu \leqslant \epsilon$; on en déduit que $|\int (u^2 - u_0^2)d\mu| \leqslant 2\epsilon + 2n\int |u - u_0|d\mu$, d'où la proposition.

BIBLIOGRAPHIE

[1] S. Banach, *Théorie des opérations linéaires* (Warszawa, 1932).

[2] G. Birkhoff, *Integration of functions with values in a Banach space*, Trans. Amer. Math. Soc., t.38 (1935), 357-378.

[3] M. Day, *Some uniformly convex spaces*, Bull. Amer. Math. Soc., t.47 (1941), 504-507.

[4] J. Dieudonné, *Sur le théorème de Lebesgue-Nikodym* (III), Annales de l'Université de Grenoble, t.23 (1947-48), 25-53.

[5] ——— *Sur le théorème de Lebesgue-Nikodym* (IV), J. Indian Math. Soc., 1951.

[6] ——— *Sur les espaces de Köthe*, Journal d'Analyse mathématique, 1951.

[7] J. Dixmier, *Sur un théorème de Banach*, Duke Math, J., t.15 (1948), 1057-1071.

[8] N. Dunford, *Uniformity in linear spaces*, Trans. Amer. Math. Soc., t.44 (1938),305-356.

[9] B. J. Pettis, *On integration in vector spaces*, Trans. Amer. Math. Soc., t.44 (1938), 277-304.

Université de Nancy

CERTAIN FOURIER TRANSFORMS OF DISTRIBUTIONS

EUGENE LUKACS AND OTTO SZÁSZ

1. Introduction. Fourier transforms of distribution functions are frequently studied in the theory of probability. In this connection they are called characteristic functions of probability distributions. It is often of interest to decide whether a given function $\varphi(t)$ can be the characteristic function of a probability distribution, that is, whether it admits the representation

$$\varphi(t) = \int_{-\infty}^{+\infty} e^{itx} dF(x). \tag{1}$$

Here $F(x)$ is a distribution function and $\varphi(t)$ its characteristic function. If $F(x)$ is absolutely continuous, $f(x) = \frac{dF}{dx}$ is called the frequency function (probability density). The frequency function $f(x)$ is non-negative and real for real x and $\int_{-\infty}^{+\infty} f(x)\, dx = 1$.

H. Cramér [2] gave a necessary and sufficient condition which a complex valued function $\varphi(t)$ of a real variable t must satisfy in order to be a characteristic function. His work is a simplification of an earlier result of Bochner [1]. Another criterion is due to Khintchine [4]. These general theorems are not easily applicable in practice. It is therefore desirable to derive conditions which are restricted to certain classes of functions but are applied more readily. Marcinkiewicz [6] and Lévy [5] derived necessary conditions for an entire function to be a characteristic function. A rather simple sufficient condition was given by Pólya [7]. In this paper another class of functions, namely the reciprocals of polynomials are studied.

The reciprocal of a polynomial with a single imaginary root is the characteristic function of the well known gamma distribution. The product of two characteristic functions is always a characteristic function [1, Satz 18] so that the reciprocal of a polynomial having only purely imaginary roots is always the characteristic function of a distribution. Employing the limit theorem of Lévy one can prove easily that the reciprocal of an entire function of genus 0 or 1, and having only imaginary roots is also a characteristic function. An analogous situation arose in a problem on the bilateral Laplace transform treated by Schoenberg [8].

In this note the following necessary condition is derived: If the reciprocal of a polynomial without multiple roots is a characteristic function then the following two conditions are satisfied:

Received December 27, 1949; in revised form February 5, 1950. The preparation of this paper was sponsored (in part) by the Office of Naval Research.

(a) The polynomial has no real roots. Its roots are located either all on the imaginary axis or in pairs $\pm b + ia$ symmetric with regard to this axis.

(b) If $b + ia$ (a, b, real; $a \neq 0$, $b \neq 0$) is a root of the polynomial then it has at least one root $i\alpha$ such that $\operatorname{sgn} \alpha = \operatorname{sgn} a$ and $|\alpha| \leqslant |a|$.

The assumption that the polynomial has no multiple roots is used only in deriving (b). These conditions indicate for example that functions like $\dfrac{1}{1+t^4}$ or $\dfrac{1}{(1+t^4)(1+t^2)}$, cannot be characteristic functions. A theorem due to Marcinkiewicz [6, theorem 3] shows this for the first of these functions but fails to reject the second.

2. Necessary conditions derived from elementary considerations. If a function $\varphi(t)$ is the characteristic function of a distribution then [3, p. 91]:

$$\varphi(0) = 1, \tag{2.11}$$

$$|\varphi(t)| \leqslant 1, \tag{2.12}$$

$$\varphi(-t) = \overline{\varphi(t)}. \tag{2.13}$$

If $\varphi(t)$ is the reciprocal of a polynomial of degree n one can write

$$\varphi(t) = \left\{\left(1 - \frac{it}{v_1}\right)\left(1 - \frac{it}{v_2}\right)\cdots\left(1 - \frac{it}{v_n}\right)\right\}^{-1}, \tag{2.2}$$

where $v_1, v_2, \ldots, v_n$ are complex numbers. The zeros of the polynomial are then given by

$$t_j = -iv_j \qquad (j = 1, 2, \ldots, n). \tag{2.21}$$

Condition (2.11) is satisfied; from condition (2.13) one obtains

$$\frac{(v_1 + it)\,(v_2 + it)\ldots(v_n + it)}{v_1\,v_2\ldots v_n} = \frac{(\bar{v}_1 + it)\,(\bar{v}_2 + it)\ldots(\bar{v}_n + it)}{\bar{v}_1\,\bar{v}_2\ldots\bar{v}_n}.$$

By arranging both sides according to powers of (it) and comparing the coefficients, it is seen that the elementary symmetric functions of the n numbers $v_1, v_2, \ldots, v_n$ are real, that is the numbers $v_1, \ldots, v_n$ are the roots of an equation of degree n with real coefficients. Consequently the v_j are either real or occur in pairs of conjugate complex numbers. Moreover, no v_j can be a purely imaginary number. The corresponding pole of the characteristic function would be real (by 2.21) which would contradict (2.12). These elementary considerations establish part (a) of our necessary conditions.

3. Auxiliary formulae. Consider first a polynomial having only one imaginary root. The corresponding frequency function is obtained by means of the inverse Fourier transform and is

$$(3.11)\qquad \frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{e^{-itx}dt}{(1-it/\alpha)^\lambda} = \begin{cases}\dfrac{\alpha^\lambda}{\Gamma(\lambda)}\,x^{\lambda-1}e^{-\alpha x} & \text{if } x>0,\\ 0 & \text{if } x<0,\end{cases}$$

and

$$(3.12)\qquad \frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{e^{-itx}dt}{(1+it/\beta)^\lambda} = \begin{cases}\dfrac{\beta^\lambda}{\Gamma(\lambda)}\,(-x)^{\lambda-1}e^{\beta x} & \text{if } x<0,\\ 0 & \text{if } x>0.\end{cases}$$

Here α and β are assumed to be real and positive, λ is a positive integer. If we put in (3.11) and (3.12) $t = \tau + \gamma$, where γ is a positive real constant, we obtain similar formulae for complex poles:

$$(3.21)\qquad \frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{e^{-itx}dt}{(1-it/v)^\lambda} = \begin{cases}\dfrac{v^\lambda}{\Gamma(\lambda)}\,x^{\lambda-1}e^{-vx} & \text{for } x>0,\\ 0 & \text{for } x<0,\end{cases}$$

$$(3.22)\qquad \frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{e^{-itx}dt}{(1+it/w)^\lambda} = \begin{cases}\dfrac{w^\lambda}{\Gamma(\lambda)}\,(-x)^{\lambda-1}e^{wx} & \text{for } x<0,\\ 0 & \text{for } x>0.\end{cases}$$

In these formulae $v=\alpha-i\gamma$ and $w = \beta + i\gamma$. The first formula is applied if the imaginary part of the pole of the integrand is negative, and the second if it is positive.

4. Functions whose Fourier transform is a polynomial without multiple roots. The formulae (3.11), (3.12), (3.21), (3.22) can be used to find the functions whose Fourier transform is the reciprocal of a polynomial satisfying the necessary conditions derived in §2. In the following only polynomials without multiple roots are considered. The zeros of the polynomials can be divided into four groups:

(i) zeros $i\beta_h$ ($h = 1, 2, \ldots, \mu$) on the positive imaginary axis ($\beta_h > 0$);

(ii) zeros $-i\alpha_j$ ($j = 1, 2, \ldots, \nu$) on the negative imaginary axis ($\alpha_j > 0$);

(iii) p symmetric pairs of complex roots in the upper half planes iw_k and $i\bar{w}_k$ where $w_k = c_k + id_k$, $c_k > 0$, $d_k > 0$ ($k = 1, 2, \ldots, p$);

(iv) n symmetric pairs of complex roots in the lower half planes $-iv_m$ and $-i\bar{v}_m$ where $v_m = a_m + ib_m$, $a_m > 0$, $b_m > 0$ ($m = 1, 2, \ldots, n$).

The function (2.2) can then be written

(4.1)

$$\varphi(t) = \left\{\prod_{j=1}^{\nu}\left(1-\frac{it}{\alpha_j}\right)\prod_{h=1}^{\mu}\left(1+\frac{it}{\beta_h}\right)\prod_{k=1}^{p}\left(1+\frac{it}{w_k}\right)\left(1+\frac{it}{\bar{w}_k}\right)\prod_{m=1}^{n}\left(1-\frac{it}{v_m}\right)\left(1-\frac{it}{\bar{v}_m}\right)\right\}^{-1}.$$

If $\varphi(t)$ is decomposed into partial fractions it is seen that

$$(4.11)\qquad \varphi(t) = \sum_{j=1}^{\nu}\frac{A_j}{1-it/\alpha_j} + \sum_{h=1}^{\mu}\frac{B_h}{1+it/\beta_h} + \sum_{k=1}^{p}\left(\frac{D_k}{1+it/w_k}+\frac{\bar{D}_k}{1+it/\bar{w}_k}\right) + \sum_{m=1}^{n}\left(\frac{C_m}{1-it/v_m}+\frac{\bar{C}_m}{1-it/\bar{v}_m}\right).$$

If $\varphi(t)$ is the Fourier transform of a distribution then the same is true for $\varphi(-t)$. It is seen from (4.1) that the zeros of $[\varphi(t)]^{-1}$ and of $[\varphi(-t)]^{-1}$ are symmetrical with respect to the real axis. This shows that it is sufficient to prove the statement for one of the half planes. In the following only the negative half plane will be considered. Applying formulæ (3.11), (3.12), (3.21), (3.22) the frequency function corresponding to $\varphi(t)$ is found to be

$$f(x) = \sum_{j=1}^{\nu} A_j a_j e^{-x a_j} + \sum_{m=1}^{n} (C_m v_m e^{-x v_m} + \bar{C}_m \bar{v}_m e^{-x \bar{v}_m}), \qquad \text{if } x > 0. \tag{4.2}$$

A similar expression is found for $f(x)$ if $x < 0$.

Introducing trigonometrical representation one has

$$v_m C_m = R_m e^{i\varphi_m} \qquad (m = 1, 2, \dots, n) \tag{4.31}$$

and

$$v_m C_m e^{-v_m x} + \bar{v}_m \bar{C}_m e^{-\bar{v}_m x} = 2R_m e^{-a_m x} \cos(\phi_m - b_m x). \tag{4.32}$$

Substituting (4.32) into (4.2) one has finally

$$f(x) = \sum_{j=1}^{\nu} A_j a_j e^{-a_j x} + 2 \sum_{m=1}^{n} R_m e^{-a_m x} \cos(\phi_m - b_m x), \qquad \text{if } x > 0. \tag{4.4}$$

5. Derivation of condition (b). We need the following lemma.

LEMMA. *A non-negative generalized trigonometric polynomial without constant term is identically zero.*

A function $g(x)$ is said to be a generalized trigonometric polynomial if

$$g(x) = \lambda_0 + \sum_{i=1}^{n} (\lambda_i \cos b_i x + \mu_i \sin b_i x), \tag{5.1}$$

where the numbers $b_1, \dots, b_n$ are arbitrary real quantities but not necessarily integers.

Proof. Let $g(x)$ be a generalized trigonometric polynomial without constant term i.e. $\lambda_0 = 0$. Since $\int_0^x g(t)\, dt$ is bounded we have

$$\lim_{x\to\infty} \frac{1}{x} \int_0^x g(t)\, dt = 0. \tag{5.2}$$

If $g(x_0) > 0$ for some x_0 then $g(x) > 2\epsilon$ for $|x - x_0| \leqslant \delta$. The function $g(x)$ is almost periodic, therefore there is an $L > 0$ such that in any interval $(y, y + L)$ a translation number τ exists such that $|g(x) - g(x + \tau)| < \epsilon$ for all x. Therefore in any interval $(y, y + L)$ there is a subinterval of length $\geqslant \delta$ where $|g(x)| \geqslant \epsilon$. But this implies

$$\lim_{x\to\infty} \frac{1}{x} \int_0^x g(t)\, dt \geqslant \frac{\epsilon\delta}{L} > 0$$

in contradiction to (5.2). We proceed now to the proof of condition (b).

Denote the pairs of roots of $[\varphi(t)]^{-1}$ in the lower half plane by $-iv_m$ and $-i\bar{v}_m$ where $v_m = a_m + ib_m$ $(m = 1, 2, \dots, n)$. If the polynomial has also ν roots

on the negative imaginary axis we denote them by $-i\alpha_j$ $(j = 1, 2, \dots \nu)$. The notation should be such that

(5.3) $$a_1 \leqslant a_2 \leqslant \dots \leqslant a_n$$

and if $\nu > 0$

(5.4) $$\alpha_1 < \alpha_2 < \dots < \alpha_\nu.$$

We observe that no term in (4.4) can vanish identically.

Let $\varphi(t)$ be a characteristic function; we assume first that $\nu > 0$ and show $\alpha_1 \leqslant a_1$. We carry the proof indirectly and suppose $a_1 < \alpha_1$. The frequency function $f(x)$ of $\varphi(t)$ as given by (4.4) is non-negative for $x > 0$. This may be written in the form

(5.5) $$\sum_{j=1}^{\nu} A_j\alpha_j e^{-\alpha_j x} + \sum_{m=1}^{n} e^{-a_m x}(\lambda_m \cos b_m x + \mu_m \sin b_m x) \geqslant 0 \quad \text{for } x > 0.$$

We define ρ by $a_1 = a_2 = \dots = a_\rho < a_{\rho+1}$ and put

$$g(x) = \sum_{m=1}^{\rho} (\lambda_m \cos b_m x + \mu_m \sin b_m x).$$

The function $g(x)$ cannot vanish identically and we have from (5.5)

(5.6) $$g(x) + \sum_{j=1}^{\nu} A_j\alpha_j e^{-\alpha'_j x} + \sum_{m=\rho+1}^{n} e^{-a'_m x}(\lambda_m \cos b_m x + \mu_m \sin b_m x) \geqslant 0 \quad \text{for } x > 0.$$

Here $\alpha'_j = \alpha_j - a_\rho$ and $a'_m = a_m - a_\rho$ and $\alpha'_j > 0$, $a'_m > 0$ on account of (5.3) and (5.4). By the Lemma, $g(x)$ must assume negative values and we see (as in the proof of the Lemma) that $g(x) < -\epsilon$ for some $\epsilon > 0$ and arbitrarily large x. But this is in contradiction to (5.6). If $\nu = 0$ i.e. if there are no roots on the imaginary axis the term $\sum_{j=1}^{\nu} A_j\alpha_j e^{-\alpha'_j x}$ is missing in (5.6). The preceding argument can nevertheless be applied to show that $[\varphi(t)]^{-1}$ must have a root on the imaginary axis.

References

[1] S. Bochner, *Monotone Funktionen, Stieltjessche Integrale und harmonische Analyse*, Math. Ann., vol. 108, (1933), 378-410.
[2] H. Cramér, *On the representation of a function by certain Fourier integrals*. Trans. Amer. Math. Soc., vol. 46, (1939), 191-201.
[3] ——— *Mathematical methods of statistics* (Princeton University Press, 1946).
[4] A. Khintchine, *Zur Kennzeichnung der charakteristischen Funktionen*. Bull. Math. Univ. Moscou, vol. 1 (1937), 1-31.
[5] P. Lévy, *L'arithmétique des lois de probabilités*, J. Math. Pures Appliquées, vol. 17 (1938), 17-39.
[6] J. Marcinkiewicz, *Sur une propriété de la loi de Gauss*, Math. Zeit., vol. 44 (1938), 612-618.
[7] G. Pólya, *Remarks on characteristic functions*, Proceedings of the Berkeley Symposium on mathematical statistics and probability (1949), 115-123.
[8] I. J. Schoenberg, *On totally positive functions, Laplace integrals and entire functions of the Laguerre-Pólya-Schur type*, Proc. Nat. Acad. Sci., vol. 33 (1947), 11-17.

Our Lady of Cincinnati College and Naval Ordinance Test Station, Inyokern
University of Cincinnati and National Bureau of Standards

AUTOMETRIZED BOOLEAN ALGEBRAS II: *THE GROUP OF MOTIONS OF B*

DAVID ELLIS

1. Introduction. The writer [1] has previously examined the fundamental concepts of distance geometry in a Boolean algebra, B, with distance defined[1] by $d(x, y) = xy' + x'y$. Any technical terms from distance geometry which are not defined in this paper may be found in [1]. A Boolean algebra bearing the given distance function is called an *autometrized Boolean algebra*. It is clear that the set of motions B (congruences of B with itself) form a group under substitution. This group we denote by $M(B)$.

M. H. Stone has shown [2] that the point set of any Boolean algebra, B, forms a ring, called the *associated Boolean ring* of B and denoted by $R(B)$, under the operations $a \oplus b = d(a, b)$, $a \otimes b = ab$, the composition of whose additive group, denoted by $G(B)$, is precisely the function $d(a, b)$.

In this paper $M(B)$ is examined more extensively than was done in [1]. It is shown that $M(B)$ and the group of automorphisms of B are subgroups of the group of complementation-preserving bi-uniform mappings of the point set of B onto itself having only the identity mapping in common. The main result is rather surprising although easily obtained, namely: the group of motions of B is isomorphic to the additive group of the associated Boolean ring of B.

2. Preliminary results.

THEOREM 1. *If f is any motion of B and $x \in B$, then $f(x') = f'(x)$.*

Proof. Since f is a motion of B, $d(x, x') = d(f(x), f(x'))$. However,[2] $d(x, x') = 1$. Hence

$$f(x)f'(x') + f'(x)f(x') = 1. \tag{1}$$

Taking meets in (1) with $f(x)$ and $f'(x)$, respectively, one obtains

$$f(x)f'(x') = f(x) \tag{2}$$

$$f'(x)f(x') = f'(x). \tag{3}$$

Complementation and use of DeMorgan's formulas in ~~[3]~~ (3) yields

$$f(x) + f'(x') = f(x). \tag{4}$$

Received January 3, 1950; presented to the American Mathematical Society, February 24, 1950.

[1]We use the following notation for the Boolean operations: $a + b$, ab, a', and $a \subset b$ denote join, meet, complement, and inclusion (in the wide sense), respectively.

[2]We denote the first and last elements of B by 0 and 1, respectively. They are, of course, also the unit elements for addition and multiplication, respectively, in the associated Boolean ring of B.

Then from (2) and (4) it is seen that $f(x) = f'(x')$ so that $f'(x) = f(x')$.

LEMMA 1. *If f is a motion of B sending a,b into $f(a), f(b)$ and $x \in B$, then*

$$f(x) = f(a)f(b)(abx + a'b'x') + f'(a)f'(b)(a'b'x + abx') + f'(a)f(b)(a'bx + ab'x') + f(a)f'(b)(ab'x + a'bx'). \tag{5}$$

Proof. This result was proved in [1].

LEMMA 2. *If f is any motion of B and $x \in B$, then*

$$f(x) = f'(0)x + f(0)x' = d(f(0), x). \tag{6}$$

Proof. One obtains (6) by setting $a=0$, $b=1$ in (5) and applying Theorem 1.

LEMMA 3. *If $d(a, c) = b$, then $d(a, b) = c$.*

Proof. This result was proved in [1].

COROLLARY. *If f is any motion of B and $x \in B$, then $d(x, f(x)) = d(0, f(0))$. Hence no motion of B other than the identity mapping leaves any point of B fixed and every motion of B is a translation in the sense that it moves each point the same distance.*

Proof. From Lemma 2, $d(f(0), x) = f(x)$ so that by Lemma 3, $f(0) = d(x, f(x))$. However, $d(0, f(0)) = f(0)$.

COROLLARY. *Although the group of motions, $M(B)$, of B and the group of automorphisms of B (B being treated as a Boolean algebra) are subgroups of the group of complementation-preserving bi-uniform mappings of the point set of B onto itself, they have only the identity mapping in common.*

Proof. Every automorphism of B leaves 0 fixed.

LEMMA 4. *If $x, y, z \in B$ then $d(x, d(y, z)) = d(z, d(x, y))$.*

Proof. This may be easily verified by direct expansion or by recalling that $d(x, y)$ is the composition of $G(B)$.

LEMMA 5. *If $a, b \in B$ there is a motion, f, of B with $b = f(a)$; that is, the group of motions of B is simply transitive.*

Proof. This result was proved in [1].

3. Isomorphism of $M(B)$ and $G(B)$.

THEOREM 2. *The groups $M(B)$ and $G(B)$ are isomorphic.*

Proof. Let $a \in B$. Then by Lemma 5, there is an $f \in M(B)$ with $f(0) = a$. This correspondence is one-to-one between $G(B)$ and $M(B)$ since:

(i). Every $a \in G(B)$ (we recall that the point sets of B and of $G(B)$ coincide) corresponds to at least one $f \in M(B)$, by Lemma 5.

(ii). No $a \in G(B)$ may correspond to more than one $f \in M(B)$, by Lemma 2 and the fact (proved in [1]) that any point forms a metric base for B.

(iii). The correspondence exhausts $M(B)$ since for $f \in M(B)$ there is an $a \in G(B)$ with $f(0) = a$.

It remains, then, only to show that if $a, b \in G(B)$ correspond to f and g, respectively, in $M(B)$, then the element of $M(B)$ corresponding to $a \oplus b$ is $f(g)$. To do this, it suffices to show that $f(g(0)) = a \oplus b$. Now

$$f(x) = d(f(0), x) = d(a, x),$$
$$g(x) = d(g(0), x) = d(b, x),$$

by Lemma 2. Hence, by Lemma 4,

$$\begin{aligned} f(g(x)) &= d(f(0), d(g(0), x)) = d(a, d(b, x)) \\ &= d(x, d(a, b)). \end{aligned}$$

Then

$$f(g(0)) = d(0, d(a, b)) = d(a, b).$$

But $d(a, b) = a \oplus b$, by definition and Theorem 2 is proved.

COROLLARY. *The group of motions of B is isomorphic to the additive group of a Boolean ring with unity and hence is an Abelian group all of whose non-zero elements have order two.*

REFERENCES

[1] David Ellis, *Autometrized Boolean algebras I*, Can. J. Math., vol. 3 (1951), 87-93.
[2] M. H. Stone, *Subsumption of Boolean algebras under the theory of rings*, Proc. Nat. Acad. Sci., vol. 20 (1934), 197-202.

University of Missouri

ON REGULAR SURFACES OF GENUS THREE

PATRICK DU VAL

ENRIQUES, in his posthumous *magnum opus* (1), devotes a chapter to the canonical or (where the genus is small) bi- or tricanonical models of regular surfaces, for various values of the genus $p_g = p_a = p$ and of the linear genus $p^{(1)}$. If, however, the cases he deals with are tabulated as follows (* marking the surfaces described by Enriques):

$p \backslash p^{(1)} =$	2	3	4	5	6	7	8	9	10
$p = 0$	*	*							
1	*	*							
2	*	*							
3		*							
4				*	*	*	*	*	
5								*	*

it is immediately clear that the scheme has remarkable gaps. The triangular space in the lower left-hand part corresponds, by the inequalities

$$p^{(1)} \geqslant 2p - 3, \qquad p^{(1)} \geqslant 3p - 6$$

(the former holding if the canonical system is irreducible, the latter if it is also simple) to a real absence of surfaces with the genera in question, except that the case $p = 5$, $p^{(1)} = 7$ has been omitted; this is easily seen to be a double normal rational ruled cubic, branching along a C^{16} (curve of order 16) which meets each generator in six and the directrix in four points; there is an analogous surface for every value of $p \geqslant 4$ satisfying the first inequality above with equality, consisting of a double normal rational ruled surface of order $p - 2$, branching along a C^{4p-4} which meets each generator in 6 points, intersection of the surface with a sextic hypersurface residual to $2p - 8$ generators. The rectangular gap in the upper right-hand part of the scheme, however, represents surfaces which presumably exist but have not been investigated; and this gap penetrates so far that the body of surfaces described is cut into two isolated parts; no surface with $p^{(1)} = 4$ is mentioned in the book, and only one with $p = 3$. It is as a first effort to fill in some of these lacunae that I offer this investigation of surfaces of genus $p = 3$, which (as might be expected) becomes less complete with increase of $p^{(1)}$.

The canonical regular surface of genus $p = 3$ and $p^{(1)} = n + 1$ is of course an n-ple plane, branching along a C^{2n+4} of some sort. The case $n = 2$ is well

Received January 23, 1950.

known, the branch curve being the most general C^8; the case $n = 3$ I studied myself some years ago (2), and showed that the branch curve has 24 cusps at the intersections of a quartic and a sextic, its equation being linearly dependent on the square of the sextic and the cube of the quartic, and is the most general curve of this kind. I also proved the case $n = 3$ of the following theorem, which I shall now prove generally, and of its converse, which I can only prove in its entirety (and which is probably only true) for $n \leqslant 5$. (We use $[n]$ to denote a projective space of n dimensions.)

THEOREM I. *In* $[n+3]$ *if* V_n^4 *is the cone projecting a Veronese surface* V_2^4 *from an* $[n-3]$, Ω, *and* U_5^n *a five dimensional manifold whose general* $[n]$ *section is the del Pezzo surface of order n, then the intersection of* V_n^4 *with* U_5^n *is in general a surface* F^{4n}, *the bicanonical model of a surface with* $p = 3$, $p^{(1)} = n+1$.

For F^{4n} has on it a net $|C|$ of curves C^{2n}, traced on it by the quadric cones Γ_{n-1}^2 of V_n^4 which project the conics of V_2^4 from the vertex Ω. Each of these lies in the $[n]$ containing the cone Γ_{n-1}^2, and is the intersection of Γ_{n-1}^2 with a del Pezzo surface S_2^n, the section of U_5^n by $[n]$. Now a quadric section of S_2^n is a canonical curve of genus $n+1$, since when the surface is mapped on a plane in the ordinary way, it appears as a sextic whose adjoint cubics are precisely the system mapping the hyperplane sections of the surface. Thus $|C'|$, the adjoint system of $|C|$, consists of the hyperplane sections of the surface F^{4n}, which are clearly the system $|2C|$, i.e.

$$|C| = |C' - C|$$

is the canonical system on the surface. Thus F^{4n} is bicanonical; it has $p = 3$ since $|C|$ is a net, and $p^{(1)} = n+1$ since this is the genus of C^{2n}.

The next problem of course is, what is the nature of the canonical n-ple plane, the projective model of $|C|$? In this connexion we prove

THEOREM II. *The canonical model of the surface* F^{4n} *is an n-ple plane, branching along a curve of order* $4n$ *with* $24(n-2)$ *cusps and* $8(n-2)(n-3)$ *nodes.*

For F^{4n} is projected from the vertex Ω_{n-3} of V_n^4 into an n-ple Veronese surface V_2^4, which when mapped on the plane in the ordinary way gives the required n-ple plane; and this n-ple V_2^4 is just the section by V_2^4 of the n-ple [5], projection of U_5^n from Ω_{n-3}; the nature of the branching of this can in turn be found by considering that of its general n-ple plane, the general projection of a del Pezzo surface of order n from $[n-3]$, whose branch curve is a C^{2n} of genus $n+1$, the projection of a general quadric section of the surface, and of class 12, since the del Pezzo surface is of class 12. From these data we see that this C^{2n} has 6 $(n-2)$ cusps and 2 $(n-2)(n-3)$ nodes; and the n-ple [5] accordingly branches along a hypersurface of order $2n$ with three dimensional cuspidal and nodal loci of orders $6(n-2)$, $2(n-2)(n-3)$ respectively; and from the intersections of these with a general V_2^4 and the mapping of the latter on the plane, Theorem II follows.

It is not likely, of course, that the branch curve is completely characterized by the mere number of its singularities. For $n = 3$ I have already given the stronger result, and for $n = 4$ it will appear in the sequel (Theorem V). It may be noted that Theorem II accords with the general formula

$$p_n = n + \tfrac{1}{2}(\beta - 1)(\beta - 2) - \delta - \kappa - 2$$

for an n-ple plane with 2β-ic branch curve having 4δ nodes and 3κ cusps.

Meanwhile, let us consider to what extent we can establish the converse of Theorem I. For a general value of n probably the best we can look for is the following:

THEOREM III. *Every regular surface with $p = 3$, $p^{(1)} = n + 1$, has as its bicanonical model a surface of order $4n$ lying on V_n^4, the canonical system being traced on it by the quadric cones Γ_{n-1}^2.*

For the characteristic series of the canonical system $|C|$ is a semicanonical g_n^1. On the canonical model of a general C^{2n}, each set of g_n^1 is joined by an $[n-2]$, and any two of these sets are together a hyperplane section of C^{2n}, i.e. any two of the ∞^1 $[n-2]$'s are joined by a hyperplane in the ambient $[n]$ of C^{2n}; and as clearly not all of these $[n-2]$'s lie in any one hyperplane, they all pass through one $[n-3]$ and generate a cone with this vertex, which is quadric since any hyperplane through one of them contains just one other. (We have made use of the dual of the familiar theorem that any set of lines, of which every two meet, either all pass through one point or all lie in one plane.)

Now on the bicanonical model of the surface there are ∞^2 sets of n points, any two of which belong to the semicanonical involution on a C^{2n}, so that the $[n-2]$'s joining them intersect in an $[n-3]$. The theorem just quoted about sets of lines every two of which meet, can easily be generalized to read: "Every set of $[k]$'s, every two of which meet in a $[k-1]$ and are joined by a $[k+1]$, either all lie in one $[k+1]$ or all pass through one $[k-1]$." Thus the ∞^2 $[n-2]$'s in the ambient F^{4n} (since they manifestly do not all lie in one $[n]$) all pass through one $[n-3]$ Ω, the common vertex of the quadric cones containing the individual curves C^{2n}; the ∞^2 $[n-2]$'s thus generate the cone V_n^4 with vertex Ω, and Theorem III is proved.

In the cases $n = 2, 3$ we can of course go further than this, and assert that the surface given by Theorem I is the most general with the given genera (for $n = 2$ this follows at once from the fact that the most general canonical surface is the double plane with octavic branch curve; for $n = 3$ it was proved in my former paper). We shall now show that the same thing is true for $n = 4, 5$. For $n = 4$ we have in fact

THEOREM IV. *The bicanonical model of the most general regular surface with $p = 3$, $p^{(1)} = 5$ is the intersection of V_4^4 with U_6^4, i.e. the complete intersection of V_4^4 with two general quadric hypersurfaces in* [7].

By Theorem III we already know that F^{16} lies on V_4^4, and its canonical system $|C|$ is traced by the quadric cones Γ_3^2. Now each C_8 of $|C|$, being a canonical curve of genus 5, is the complete intersection of a net of quadrics in its ambient [4], one of which is of course Γ_3^2. The rest trace on each generating plane of Γ_3^2 a pencil of conics (whose base points are the four points in which the plane meets F^{16}), on Γ_3^2 itself a pencil of Segre (quartic del Pezzo) surfaces S_2^4 (whose base curve is C^8), and on the vertex line Ω of V_4^4 an involution. Thus (since any two generating planes of V_4^4 belong to a Γ_3^2) the pencils of conics in all these planes with base points at the intersections with F^{16} trace the same involution on Ω; and isolating in each pencil the conic tracing a particular pair of the involution, we see that the locus of these is a three-dimensional variety on V_4^4, which traces on each generating plane a conic, and on each Γ_3^2 an S_2^4; this variety does not contain Ω, but meets it in a pair of points, and must accordingly be a quadric section, since every $(n-1)$ dimensional variety on V_n^4 is either a complete intersection or residual to a Γ_{n-1}^2, and in the latter case must contain Ω_{n-3}; and it contains F^{16}. Thus the ∞^1 pairs of the involution on Ω give a pencil of quadric sections of V_4^4, all containing F^{16}, which is thus the complete intersection of V_4^4 with a pencil of quadric hypersurfaces in its ambient [7], i.e. with their base U_6^4. Theorem IV is thus proved.

Since V_4^4 is itself the base of an ∞^5 linear system of quadrics, the pencil whose intersection is U_6^4 is not unique, but is an arbitrary pencil skew to the ∞^5 system within an ∞^7 linear system containing the latter. F^{16} thus lies on ∞^{12} U_6^4's.

Theorem IV enables us to specify more precisely the branch curve of the canonical quadruple plane as follows:

THEOREM V. *The branch curve of the canonical surface with $p = 3$, $p^{(1)} = 5$, is the envelope of a family of quartic curves of index 3 (i.e. depending cubically on a parameter, so that three curves of the family pass through a general point of the plane) of which two members reduce to double conics, the eight branch points on each of these being the 16 nodes of the branch curve. Conversely the envelope of the most general family of this kind is the branch curve of a quadruple plane, which is a canonical surface with $p = 3$, $p^{(1)} = 5$.*

For the discriminant of the pencil of conics traced by the pencil of quadrics Q_6^2 on any plane through Ω can be written in the form

$$\alpha\lambda^3 + 3\beta\lambda^2 + 3\gamma\lambda + \delta$$

where λ is the parameter in the pencil, and $\alpha, \beta, \gamma, \delta$ are quadratic functions of the coordinates of the plane, i.e. of those of the point in which it meets the [5] onto which we project. Thus the branch hypersurface of the quadruple [5] is the envelope of the family of quadrics

$$\alpha\lambda^3 + 3\beta\lambda^2 + 3\gamma\lambda + \delta = 0 \tag{†}$$

and its cuspidal locus is the complete intersection of the three quartics

$$\alpha\gamma = \beta^2, \ \alpha\delta = \beta\gamma, \ \beta\delta = \gamma^2.$$

For the two values of λ, for which Q_6^2 touches Ω, the Q_4^2 given by (†) reduces to a double [4], branching along a Q_3^2, and these two Q_3^2's together constitute the nodal locus of the branch hypersurface, since the plane joining Ω to a point of either of them meets the corresponding Q_6^2 in a double line, and hence F^{16} in four points which coincide by pairs. That this branch hyperplane is the envelope of the most general family of quadrics answering to this description follows from the fact that its general hyperplane section, branch curve of the projected Segre surface, is the envelope of the most general family of conics, of index 3, of which two members reduce to double lines. Such a family can in fact be represented by an equation of the form

$$ax^2\lambda^3 + 3\beta\lambda^2 + 3\gamma\lambda + dy^2 = 0$$

which contains 14 homogeneously entering coefficients; and bearing in mind the ∞^4 projective transformations which leave the x and y axes invariant, and the possibility of multiplying the parameter λ by a constant, we see that the number of such envelopes projectively distinct is ∞^8; but the number of projectively distinct figures in [4] consisting of a Segre surface and a line to project it from is also ∞^8, and both systems are irreducible. Since the branch locus of the quadruple V_2^4 is the section of that of the quadruple [5], Theorem V follows.

Turning now to the case $n = 5$ we have

THEOREM VI. *The bicanonical model of the most general regular surface with* $p = 3$, $p^{(1)} = 6$ *is the intersection of* V_5^4 *with* U_5^5.

Up to a point the proof of this is very parallel to that of Theorem IV. The most general canonical curve of genus 6 lies on precisely one del Pezzo surface of order 5, S_2^5, of which it is a quadric section. Each C^{10} of the canonical system $|C|$ on F^{20} is thus the complete intersection of the cone Γ_4^2 with a determinate S_2^5, which of course is itself the complete intersection of a linear system of ∞^4 quadrics Q_4^2; these trace on each generating [3] of Γ_4^2 the ∞^4 quadric surfaces through the five intersections of the [3] with F^{20}, and on the vertex plane Ω a linear system of ∞^4 conics. Thus just as in Theorem IV, in all the ∞^2 generating [3]'s of V_5^4, the quadric surfaces through the intersections of [3] with F^{20} trace the same system of conics in Ω; and the locus of the quadric surface in each of these systems which traces a particular conic in Ω is a quadric section of V_5^4. It is not, of course, the section by a determinate quadric, since V_5^4 is itself the intersection of ∞^5 quadrics; but in the ambient [5], Λ_5 say, of C^{10}, the quadric Q_4^2 containing the S_2^5 on which C^{10} lies and tracing the chosen conic on Ω is determinate, and the locus of these ∞^2 Q_4^2's is clearly a quadric section of the cubic hypersurface K_7^3 generated by the ∞^2 Λ_5's, the cone projecting from Ω the cubic symmetroid generated by the conic planes of V_2^4; it is the section of K_7^3 moreover by a determinate quadric, since of course no quadrics

contain K_7^3. We have thus in the ambient [8] of the whole figure a linear system of ∞^4 quadric hypersurfaces Q_7^2, whose complete intersection with K_7^3 is the locus Σ_4 of the surfaces S_2^5 in the ∞^2 Λ_5's, and with V_5^4 is the surface F^{20}.

The hypersurface K_7^3 is also generated of course by the ∞^2 tangent [5]'s M_5 of V_5^4, each of which touches it all over a generating [3], and two of which pass through a general generating [3] of K_7^3. Each M_5 is met by ∞^1 Λ_5's in the pencil of [4]'s through its [3] of contact, and accordingly meets Σ_4 in the surface generated by the elliptic quintic curves traced by these [4]'s on the S_2^5's in the corresponding Λ_5's, which is clearly itself an S_2^5, since the curves all meet the [3] of contact in the same five points, the intersections of this [3] with F^{20}. Since K_7^3 has [6] sections consisting of a repeated Λ_5 and an M_5, Σ_4 is of order 15.

We shall now show that the complete intersection of the ∞^4 quadrics Q_7^2 is five dimensional. If it were of more dimensions, its intersection with K_7^3 would be more than four dimensional, which it is not. If it were less, it must consist of Σ_4^{15}, together possibly with some residual variety R which may be two, three, or four dimensional; and any [5] which does not lie on any of the quadrics Q_7^2 must in this case meet them in a linear system of quadrics Q_4^2 whose complete intersection is precisely the section by [5] of $\Sigma_4^{15}+R$, i.e., a curve. Now consider the [5] X_5, joining a generating [3] of K_7^3 to the [3]'s of contact of the two M_5's which intersect in this [3]. (This join is in fact a [5], since the three [3]'s all pass through the plane Ω and do not belong to a pencil.) X_5 meets K_7^3 in three [4]'s, joining the three [3]'s by pairs, and of these, two lie in M_5's and one in a Λ_5, so that each of them meets Σ_4^{15} in an elliptic quintic curve; and these three quintic curves, of which each pair has in common the five points traced by the [3] in which their ambient [4]'s intersect, are the complete intersection of Σ_4^{15} with X_5. Now the number of projectively distinct figures in [5] consisting of three normal elliptic quintic curves each pair of which have a [3] section in common is ∞^{15}, which is also the number of projectively distinct figures consisting of an S_2^5 and three hyperplanes, and both systems are clearly irreducible; thus the three curves just obtained are the sections by their ambient [4]'s of an S_2^5 lying in X_5, the ∞^4 quadrics through which are precisely the sections by X_5 of the ∞^4 quadrics Q_7^2, since a quadric in any space is completely determined when three hyperplane sections are given. Thus the complete intersection of the quadrics Q_7^2 meets X_5 not in a curve but in a surface, and is accordingly five dimensional. Since moreover it is met by the ∞^2 [5]'s Λ_5, the ∞^2 [5]'s M_5 and the ∞^4 [5]'s X_5 in del Pezzo surfaces S_2^5, it must be U_5^5, and F^{20} is its complete intersection with V_5^4; Theorem VI is thus proved.

It is to be noted that there are four descriptively different types of U_5^5; first what we shall regard as the general case, the Grassmannian of a linear complex of lines in [4]; and the cones projecting from a point, line, and plane respectively a [7], [6], and [5] section of this. The last three, however, are all special or limiting forms of the first, although they do not occur among the hyperplane sections of the general U_6^5, Grassmannian of all the lines of [4]; since the equations of the five linearly independent quadrics Q_7^2 can be taken to be

$$x_{ij}x_{kl} + x_{ik}x_{lj} + x_{il}x_{jk} = 0,$$

where i, j, k, l are any four of 1, 2, 3, 4, 5, and $x_{ij} = -x_{ji}$ are any ten linear functions of the coordinates, the four cases occurring according as these ten functions satisfy one, two, three, or four independent linear identities.

I have not succeeded in this case ($n = 5$) in finding any more precise specification of the branch curve of the canonical quintuple plane (such as was provided for $n = 4$ by Theorem V, and for $n = 3$ is to be found in my former paper) than is given by Theorem II.

It does not seem very probable that for $n = p^{(1)} - 1 \geqslant 6$ the surface F^{4n} of Theorem I is the most general of its genera, since for $n \geqslant 6$ the quadric section of the del Pezzo surface is not the most general canonical curve of its genus, and it is therefore not likely that the section of this surface by a quadric cone with $[n - 3]$ vertex is the most general such curve with a semicanonical g_n^1, a point which was essential in the proofs of Theorems IV, VI, and the analogous result for $n = 3$. In any case, for $n \geqslant 10$ the surface given by Theorem I does not exist, since there is no del Pezzo surface; and if in Theorem I we interpret U_3^n to include as a special case the cone projecting a normal elliptic cone from [3], F^{4n} acquires four elliptic conical nodes, intersections of this [3] with V_n^4, which reduce its arithmetic genus to -1, and is in fact equivalent to an elliptic ruled surface, being generated by an elliptic pencil of rational quartics. The case $n = 6$ ($p^{(1)} = 7$) is perhaps crucial, in the sense that if the analogue of Theorems IV, VI could be established in this case, by some other method than that used hitherto, it might be a plausible speculation whether it held for all values of n; there is of course no reason *a priori* why the canonical curves on a surface of given genera should be the most general compatible with the existence of the semicanonical series required—though I believe this is so in all the cases that have been studied.

REFERENCES

(1) F. Enriques, *Le superficie algebriche* (Bologna, 1949), capitolo VIII.

(2) P. Du Val, *On the regular surface of genus three and linear genus four*, J. Lond. Math. Soc., vol. 8 (1933), 11-16.

University of Georgia

A CYCLIC INVOLUTION OF PERIOD ELEVEN

W. R. HUTCHERSON

In two earlier papers* the writer discussed involutions of periods five and seven on certain cubic surfaces in S_3. In this paper, a quartic surface containing a cyclic involution of period eleven is considered.

The surface

$$F_4(x_1,x_2,x_3,x_4) \equiv ax_2x_3^3 + bx_1x_2x_4^2 + cx_1x_3^2x_4 + dx_2^2x_3x_4 = 0$$

is invariant under the cyclic collineation T of period eleven,

$$x'_1:x'_2:x'_3:x'_4 = x_1:Ex_2:E^2x_3:E^3x_4 \qquad (E^{11} = 1).$$

Points $P_1(1,0,0,0)$, $P_2(0,1,0,0)$, $P_3(0,0,1,0)$, and $P_4(0,0,0,1)$ are all invariant under T and lie on the surface F_4. This fact may be stated in the following theorem.

THEOREM 1. *Each vertex of the tetrahedron of reference not only lies on the surface but is a point of coincidence.*

By rewriting F_4 in the order

$$ax_2x_3^3 + x_4(bx_1x_2x_4 + cx_1x_3^2 + dx_2^2x_3) = 0$$

it is easily seen that the line P_1P_2 ($x_3 = x_4 = 0$) lies on the surface. However, only the two points P_1 and P_2 of the line are invariant under T. In similar manner P_1P_4, P_1P_3, P_2P_4, and P_3P_4 lie on F_4 with only two invariant points on each line. The line P_2P_3 does not lie on the surface. A second theorem has been proved.

THEOREM 2. *This surface includes all the six edges of the tetrahedron of reference, except P_2P_3.*

It is true that P_3 is simple on F_4 while P_2 and P_4 are double, and P_1 is triple. In this paper only point P_3 will be investigated in detail.

Consider a curve C, not transformed into itself by T, and passing through P_3. Take the plane $x_4 + Kx_1 = 0$ of the pencil passing through P_2 and P_3, tangent to C. This plane is transformed into $E^3x_4 + Kx_1 = 0$ or $x_4 + KE^8x_1 = 0$ by T and hence is non-invariant. The curve cut out on F_4 by $x_4+Kx_1=0$ is therefore non-invariant. The common tangent to the two curves is not

Received September 26, 1949. Presented to the American Mathematical Society, September 1, 1949.

*W. R. Hutcherson, *Maps of certain cyclic involutions on two-dimensional carriers*, Bull. Amer. Math. Soc., vol. 37 (1931), 759-765; *A cyclic involution of order seven*, Bull. Amer. Math. Soc., vol. 40 (1934), 143-151.

transformed into itself. Thus the two curves do not touch each other at P_3. Now, since C was a variable curve through P_3 satisfying the non-invariant property, it follows that P_3 is an imperfect coincidence point. In similar manner it can be shown that P_1, P_2, and P_4 are also imperfect coincidence points. The following theorem has just been proved.

THEOREM 3. *The I_{11} belonging to F_4 in S_3 has four imperfect points of coincidence.*

Consider the complete system of curves $|A|$ cut out on F_4 by all surfaces of order eleven. Its dimension is 243, its genus is 243, and the number of variable intersections of two members of the system is 484. A curve A of this system is not in general transformed into itself by T. There are, however, eleven partial systems $|A_i|$ in $|A|$ which are transformed into themselves. By use of $|A_1|$ we find

$$\begin{aligned}
&a_1x_1^{11} + a_2x_2^{11} + a_3x_3^{11} + a_4x_4^{11} + a_5x_1^7x_3x_4^3 + a_6x_1^6x_2^2x_4^3 + a_7x_1^6x_2x_3^2x_4^2 \\
&+ a_8x_1^5x_2^3x_3x_4^2 + a_9x_1^4x_2^5x_4^2 + a_{10}x_1^6x_3^4x_4 + a_{11}x_1^5x_2^2x_3^3x_4 + a_{12}x_1^4x_2^4x_3^2x_4 \\
&+ a_{13}x_1^3x_2^6x_3x_4 + a_{14}x_1^2x_2^8x_4 + a_{15}x_1^5x_2x_3^5 + a_{16}x_1^4x_2^3x_3^4 + a_{17}x_1^3x_2^5x_3^3 \\
&+ a_{18}x_1^2x_2^7x_3^2 + a_{19}x_1x_2^9x_3 + a_{20}x_1^3x_2x_4^7 + a_{21}x_1^3x_3^2x_4^6 + a_{22}x_1^2x_2^2x_3x_4^6 \\
&+ a_{23}x_1x_2^4x_4^6 + a_{24}x_1^2x_2x_3^3x_4^5 + a_{25}x_1x_2^3x_3^2x_4^5 + a_{26}x_2^5x_3x_4^5 + a_{27}x_1^2x_3^5x_4^4 \\
&+ a_{28}x_1x_2^2x_3^4x_4^4 + a_{29}x_2^4x_3^3x_4^4 + a_{30}x_1x_2x_3^6x_4^3 + a_{31}x_2^3x_3^5x_4^3 + a_{32}x_1x_3^8x_4^2 \\
&+ a_{33}x_2^2x_3^7x_4^2 = 0.
\end{aligned}$$

We refer the curves A_1 projectively to the hyperplanes of a linear space of thirty-two dimensions. We obtain a surface φ, of order 44, as the image of I_{11}. The equations of the transformation for mapping I_{11} upon φ in S_{32} are

$$\begin{array}{lll}
\rho X_1 = x_1^{11} & \rho X_{12} = x_1^4x_2^4x_3^2x_4 & \rho X_{23} = x_1x_2^4x_4^6 \\
\rho X_2 = x_2^{11} & \rho X_{13} = x_1^3x_2^6x_3x_4 & \rho X_{24} = x_1^2x_2x_3^3x_4^5 \\
\rho X_3 = x_3^{11} & \rho X_{14} = x_1^2x_2^8x_4 & \rho X_{25} = x_1x_2^3x_3^2x_4^5 \\
\rho X_4 = x_4^{11} & \rho X_{15} = x_1^5x_2x_3^5 & \rho X_{26} = x_2^5x_3x_4^5 \\
\rho X_5 = x_1^7x_3x_4^3 & \rho X_{16} = x_1^4x_2^3x_3^4 & \rho X_{27} = x_1^2x_3^5x_4^4 \\
\rho X_6 = x_1^6x_2^2x_4^3 & \rho X_{17} = x_1^3x_2^5x_3^3 & \rho X_{28} = x_1x_2^2x_3^4x_4^4 \\
\rho X_7 = x_1^6x_2x_3^2x_4^2 & \rho X_{18} = x_1^2x_2^7x_3^2 & \rho X_{29} = x_2^4x_3^3x_4^4 \\
\rho X_8 = x_1^5x_2^3x_3x_4^2 & \rho X_{19} = x_1x_2^9x_3 & \rho X_{30} = x_1x_2x_3^6x_4^3 \\
\rho X_9 = x_1^4x_2^5x_4^2 & \rho X_{20} = x_1^3x_2x_4^7 & \rho X_{31} = x_2^3x_3^5x_4^3 \\
\rho X_{10} = x_1^6x_3^4x_4 & \rho X_{21} = x_1^3x_3^2x_4^6 & \rho X_{32} = x_1x_3^8x_4^2 \\
\rho X_{11} = x_1^5x_2^2x_3^3x_4 & \rho X_{22} = x_1^2x_2^2x_3x_4^6 & \rho X_{33} = x_2^2x_3^7x_4^2
\end{array}$$

By eliminating ρ, x_1, x_2, x_3, and x_4 from these thirty-three equations and $F_4(x_1x_2x_3x_4) = 0$, we get as the thirty equations defining the surface:

$$\left\| \begin{matrix} X_1 & X_5 & X_6 & X_7 & X_8 & X_9 & X_{11} & X_{12} & X_{16} \\ X_5 & X_{21} & X_{22} & X_{24} & X_{25} & X_{26} & X_{28} & X_{29} & X_{31} \end{matrix} \right\| = 0$$

$$\left\| \begin{matrix} X_2 & X_9 & X_{13} & X_{14} & X_{17} & X_{18} & X_{19} \\ X_{13} & X_5 & X_7 & X_8 & X_{10} & X_{11} & X_{12} \end{matrix} \right\| = 0$$

$$\left\|\begin{matrix} X_3 & X_{15} & X_{27} & X_{30} & X_{31} & X_{32} & X_{33} \\ X_{10} & X_6 & X_{20} & X_{22} & X_{23} & X_{24} & X_{25} \end{matrix}\right\| = 0$$

$$\left\|\begin{matrix} X_4 & X_{20} & X_{22} & X_{23} & X_{24} & X_{25} \\ X_{21} & X_7 & X_{11} & X_{12} & X_{15} & X_{16} \end{matrix}\right\| = 0$$

$$\left\|\begin{matrix} X_6 & X_7 & X_8 & X_{10} & X_{11} \\ X_{23} & X_{25} & X_{26} & X_{28} & X_{29} \end{matrix}\right\| = 0$$

and equation $aX_{31} + bX_{25} + cX_{23} + dX_{29} = 0$. Designate by P'_3 the branch point of φ corresponding to the point P_3 on F_4. The coordinates of P'_3 are all zero except X_3.

The curves A_1 on F_4 pass through P_3 if $a_3 = 0$. The tangent plane at P_3 to F_4 is $x_2 = 0$. Now, the system of eleventh-degree surfaces passing through P_3 cuts $x_2 = 0$ in the curves $x_2 = 0$, and

$$a_1x_1^{11}+a_4x_4^{11}+a_6x_1^7x_3x_4^3+a_{10}x_1^6x_3^4x_4+a_{21}x_1^3x_3^2x_4^6+a_{27}x_1^2x_3^5x_4^4+a_{32}x_1x_3^8x_4^2=0.$$

For general values of the constants this is an eleventh-degree curve with a triple point at P_3, two branches being tangent to the line $x_2 = x_4 = 0$ and one to the line $x_2 = x_1 = 0$. When $a_6 = a_{10} = a_{21} = a_{27} = a_{32} = 0$, the plane eleventh-degree curve breaks up into eleven lines through P_3. These are all distinct except when either $a_1 = 0$ or $a_4 = 0$, when they coincide with $x_2 = x_4 = 0$ or $x_2 = x_1 = 0$, respectively. Since P_3 is imperfect, the $|A_1|$ through P_3 must have eleven distinct branches unless each branch touches one of the two invariant directions. In the plane $x_2 = 0$, the involution I_{11} is generated by the homography T_1, which is $x'_1 : x'_3 : x'_4 = x_1 : E^2x_3 : E^3x_4$.

By use of the plane quadratic transformation X, $y_1 : y_3 : y_4 = w_1w_4 : w_3^2 : w_1w_3$ and X^{-1}, $w_1 : w_3 : w_4 = y_4^2 : y_3y_4 : y_1y_3$ one gets

$$(w_1,w_3,w_4) \sim_{X^{-1}} (y_4^2,y_3y_4,y_1y_3) \sim_{T_1} (E^6y_4,E^5y_3y_4,E^2y_1y_3) \sim_X (E^6w_1,E^5w_3,E^2w_4)$$

or

$$x'_1 : x'_3 : x'_4 = E^4x_1 : E^3x_3 : x_4 \qquad \text{for } T_2.$$

Again $(w_1,w_3,w_4) \sim_{X^{-1}} (y_4^2,y_3y_4,y_1y_3) \sim_{T_2} (y_4^2,E^3y_3y_4,E^7y_1y_3) \sim_X (w_1,E^3w_3,E^7w_4)$ or T_3 is $x'_1 : x'_3 : x'_4 = x_1 : E^3x_3 : E^7x_4$. By use of XT_3X^{-1} one gets

$$(w_1, w_3, w_4) \sim (E^{14}w_1, E^{10}w_3, E^3w_4)$$

or T_4 is $x'_1 : x'_3 : x'_4 = E^{11}x_1 : E^7x_3 : x_4 = x_1 : E^7x_3 : x_4$.

Thus, the following theorem has just been established.

THEOREM 4. *The imperfect point of coincidence P_3 has an imperfect point in the first order neighbourhood along the $x_1 = x_2 = 0$ direction. It also has an imperfect point in the second order neighbourhood. In the third order neighbourhood there is a perfect point.*

Now, investigate the characteristics of the point adjacent to P_3 along the invariant direction $x_4 = x_2 = 0$. By use of YT_1Y^{-1}, where the transforma-

tion Y is $y_1 : y_3 : y_4 = w_3w_4 : w_3^2 : w_1w_4$ and the inverse is $w_1 : w_3 : w_4 = y_3y_4 : y_1y_3 : y_1^2$, we get $(w_1,w_3,w_4) \sim_{Y^{-1}} (y_3y_4, y_1y_3, y_1^2) \sim_{T_1} (E^5y_3y_4, E^2y_1y_3, y_1^2) \sim_Y (E^5w_1, E^2w_3, w_4)$. We have an imperfect point. Define T'_2 as YT_1Y^{-1}. Now apply $XT'_2X^{-1} \equiv T''_2$ to our next order point, remembering that T'_2 may be written $x'_1 : x'_3 : x'_4 = E^5x_1 : E^2x_3 : x_4$. We obtain

$$(w_1,w_3,w_4) \sim_{X^{-1}} (y_4^2, y_3y_4, y_1y_3) \sim_{T'_2} (y_4^2, E^2y_3y_4, E^7y_1y_3) \sim_X (w_1, E^2w_3, E^7w_4).$$

This transformation T''_2 or $x'_1 : x'_3 : x'_4 = x_1 : E^2x_3 : E^7x_4$ gives evidence of another imperfect point. For the third order neighbourhood, we use $YT''_2Y^{-1} \equiv T'''_2$. This becomes $(w_1,w_3,w_4) \sim (E^9w_1, E^2w_3, w_4)$, denoting an imperfect point in the third order neighbourhood of P_3 along the $x_2 = x_4 = 0$ direction.

Finally, by use of $XT'''_2X^{-1} \equiv T^{iv}_2$ we get $(w_1, w_3, w_4) \sim (w_1, E^2w_3, E^{11}w_4)$ or (w_1, E^2w_3, w_4) since $E^{11} = 1$. This indicates a perfect point. We shall state our result in the following theorem.

THEOREM 5. *Along the invariant direction $x_2 = x_4 = 0$, there are no perfect points in either the first or second or third order neighbourhood of P_3. There is, however, a perfect point in the fourth order neighbourhood.*

The following theorem is self-evident.

THEOREM 6. *The imperfect point P_3 on F_4 has no perfect points in the neighbourhood of the first or second order. It does have one in the third order neighbourhood and one in the fourth order neighbourhood, however.*

University of Florida

THE DUALITY THEOREM FOR CURVES OF ORDER n IN n-SPACE

DOUGLAS DERRY

LET C_n be a curve in real projective n-space which is a continuous $1-1$ image of either the projective line or one of its closed segments. Consequently its points depend continuously on a real variable s for which $0 \leqslant s \leqslant 1$, with the understanding that $s = 0$ and $s = 1$ represent the same curve point in the case that C_n is the image of the complete projective line. The points of C_n will be described by their corresponding real numbers s.

We assume

(1) No $(n-1)$-dimensional hyperplane H cuts C_n in more than n points. An immediate consequence of the above is that any $k+1$ distinct curve points generate a linear k-subspace.

We assume

(2) The linear k-subspace L generated by $k+1$ curve points always converges to a linear k-subspace designated by (k, s) as the $k+1$ points all converge to s, $0 \leqslant k < n$.

The subspaces (k, s) enable us to count multiple intersection points of a linear subspace L with C_n. A point s is said to be within L k-fold if $(k-1, s) \subset L$, $(k, s) \not\subset L$. We now assume that (1) and (2) are both true when the multiple intersection points of both H and L are counted by the above convention.

In 1936 Scherk[1] gave the first proof that the dual of C_n has properties (1) and (2). His proof first derives the result for the case where C_n is the map of the whole projective line and then derives the general result by showing that every C_n is part of such a curve. In the following an alternative proof is given which applies directly to any C_n. The methods are elementary. Use is made of the easily established fact that the projection of a C_n from one of its points s' is a C_{n-1} and each (k, s) of C_n projects either into a (k, s), $0 \leqslant k \leqslant n-2$, or into a $(k-1, s)$, $1 \leqslant k \leqslant n-1$, for the projected curve according as either $s' \neq s$ or $s' = s$.

THEOREM 1. *Where $\bar{s}$ is an interior point of C_n let s^ν_1, s^ν_2 be two sequences of real numbers which approach $\bar{s}$ and for which $s^\nu_1 \neq s^\nu_2$. If P^ν be a convergent sequence of space points selected from the intersection of $(n-1, s^\nu_1)$ and $(n-1, s^\nu_2)$ then it converges to a point P of $(n-2, \bar{s})$.*

For the proof of this result we shall use

Received February 13, 1950.

[1]P. Scherk. *Über differenzierbare Kurven und Bögen* II. Časopis pro pěstování, matematiky a fysiky 66 (1937), 172-191.

Lemma 1. *If $\bar{s}$ is an interior point of C_n and $P \in (n-1, \bar{s})$ but P* non $\in$ *$(n-2, \bar{s})$ then for every sufficiently small curve neighborhood $I(\bar{s})$ a curve neighborhood $J(\bar{s})$, $J(\bar{s}) \subset I(\bar{s})$, together with a space neighborhood $N(P)$ of P exists with the following properties*:

(1) Curve points $s, s_1, s_2, \ldots, s_{n-2}$ from $J(\bar{s})$ and a point P' of $N(P)$ build a hyperplane which cuts $I(\bar{s})$ in exactly one additional point $q(s)$. (Some or all of $s_1, s_2, \ldots, s_{n-2}$ may coincide.)

(2) As s moves continuously in one direction in $J(\bar{s})$, $q(s)$ moves continuously in the opposite direction so that $q(s') \neq q(s'')$ if $s' \neq s''$.

Proof of Lemma. As the lemma deals with local properties of C_n it is sufficient to prove it within an affine n-subspace of the projective space which contains P and $\bar{s}$. By hypothesis the linear $n-2$-subspace generated by any $n-1$ curve points will approach $(n-2, \bar{s})$ as these points all approach $\bar{s}$. Therefore and because P non $\in (n-2, \bar{s})$ a curve neighborhood $I(\bar{s})$, i.e. a set of points s containing $\bar{s}$ for which $s_a < s < s_b$, together with a point P' sufficiently close to P will always generate a hyperplane H. H converges to $(n-1, \bar{s})$ as $P' \to P$ and $s, s_1, s_2, \ldots, s_{n-2}$ converge to $\bar{s}$. The endpoints s_a, s_b of $I(\bar{s})$ will be on the same or opposite sides of H according as they are on the same or opposite sides of $(n-1, \bar{s})$ provided $s, s_1, s_2, \ldots, s_{n-2}$ are in a sufficiently small neighborhood $I'(\bar{s})$ and P' in a sufficiently small neighborhood N' of P. In this event the number of intersection points of H and $I(\bar{s})$ will be odd or even according as n is odd or even. Therefore H cuts $I(\bar{s})$ in a point $q(s)$ in addition to the points $s, s_1, \ldots, s_{n-2}$ and in no further points because of the order of C_n by (1). For fixed $s_1, s_2, \ldots, s_{n-2}$, $q(s)$ moves continuously with s because H moves continuously with s. As $q(s), s_1, \ldots, s_{n-2}$ and P' define H completely, two different positions of s cannot define the same $q(s)$ because the order of the curve would exceed n in this case. For the same reason $q(s)$ cannot experience a reversal as s moves continuously in a fixed direction. As $H \to (n-1, \bar{s})$, $q(s) \to \bar{s}$. Hence neighborhoods $J(\bar{s})$, $N(P)$ with $J(\bar{s}) \subset I'(\bar{s})$, $N(P) \subset N'$ exist so that if $s, s_1, s_2, \ldots, s_{n-2} \in J(\bar{s})$, $P' \in N(P)$ then $q(s) \in I'(\bar{s})$ Consequently $q(q(s))$ is defined and must be equal to s as $q(s), s_1, s_2, \ldots, s_{n-2}$ and P' define a unique hyperplane. If we project from $s_1, s_2, \ldots, s_{n-2}, P'$ then C_n will be projected into a curve of order two on the affine line. Points for which $s = q(s)$ will be projected into the reversal points of such a curve and as there are at most two such points we conclude $q(s) \neq s$ with at most two possible exceptions. Let $s' \in J(\bar{s})$, $q(s') \neq s'$. Then $q(s') \in I'(\bar{s})$. Let s move continuously in a fixed direction in $I'(\bar{s})$ from s' to $q(s')$. $q(s)$ will move from $q(s')$ to s' in a fixed direction and remain in $I(\bar{s})$. As $I(\bar{s})$ is not the whole curve C_n this can only happen if $q(s)$ moves in the direction opposite to that of s. The lemma is now completely proved.

We write $q(s)$ as $q(s, s_1, s_2, \ldots, s_{n-2})$ because it is a function of the $n-1$ variables $s, s_1, s_2, \ldots, s_{n-2}$. If any one of these variables moves in a fixed direction in $J(\bar{s})$ while all the others remain fixed, $q(s, s_1, \ldots, s_{n-2})$ will move

in the opposite direction. To prove the theorem we note that, as P is the limit of P^μ, $P \in (n-1, \bar{s})$. We assume P non $\in (n-2, \bar{s})$, construct neighborhoods $I(\bar{s})$, $J(\bar{s})$, $N(P)$, satisfying the conditions of the lemma and select $s^\mu{}_1, s^\mu{}_2 \in J(\bar{s})$, $P^\mu \in N(P)$. Because $P^\mu \in (n-1, s^\mu{}_1)$, $q(s^\mu{}_1, s^\mu{}_1, \ldots, s^\mu{}_1) = s^\mu{}_1$. Now if we move each of the variables successively from $s^\mu{}_1$ to $s^\mu{}_2$ the point q will move in the opposite direction and remain on $I(\bar{s})$ in accordance with the lemma. But as $I(\bar{s})$ is not the whole curve C_n and $q(s^\mu{}_2, s^\mu{}_2, \ldots, s^\mu{}_2) = s^\mu{}_2$, this is impossible. Hence $P \in (n-2, \bar{s})$ and the theorem is proved.

THEOREM 2. *If s belongs to an arc $s_1 < s < s_2$ then not all of $(n-1, s)$ can pass through a single point.*

Proof. The result is true for a C_1 as by definition two different values of s define different curve points $(0, s)$. We assume the result true for C_{n-1} and proceed by induction. Should an arc $s_1 < s < s_2$ of C_n exist together with a point P so that all $(n-1, s)$, $s_1 < s < s_2$, pass through P then by Theorem 1 all $(n-2, s)$, $s_1 < s < s_2$, must pass through the same point. If we project the curve C_n from one of its points the resulting curve is a C_{n-1} for which all $(n-2, s)$, $s_1 < s < s_2$ pass through the projection of P. This contradicts the induction assumption and thus the theorem is proved.

DEFINITION. A system of linear subspaces $S^\mu{}_r$ is defined to converge to a subspace S_r if a basis $\mathbf{a}^\mu{}_1, \mathbf{a}^\mu{}_2, \ldots, \mathbf{a}^\mu{}_{r+1}$ exists for each $S^\mu{}_r$, with $\mu \geqslant \mu_0$, such that $\mathbf{a}^\mu{}_k$, $1 \leqslant k \leqslant r+1$, converges to $\mathbf{a}_k$ where $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{r+1}$ is a basis of S_r.

LEMMA 2. *$S^\mu{}_r$ is a set of linear subspaces of dimension $\geqslant r$, $0 \leqslant r < n$, defined for positive integers μ. The limit points of any point set P^μ, $P^\mu \in S^\mu{}_r$, are all within a linear r-subspace S_r. Then $S^\mu{}_r$ converges to S_r as μ approaches infinity.*

Proof. Let T_{n-r-1} be any linear $(n-r-1)$-subspace such that the projective n-space is the direct sum of T_{n-r-1} and S_r. We choose μ_0 so large that $S^\mu{}_r$ contains no elements of T_{n-r-1} for $\mu \geqslant \mu_0$. This is possible as T_{n-r-1} is a closed compact set which contains no elements of S_r. If vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{r+1}$ form a basis of S_r each $S^\mu{}_r$, $\mu \geqslant \mu_0$ will have a basis $\mathbf{a}_1 + \mathbf{p}_1$, $\mathbf{a}_2 + \mathbf{p}_2, \ldots, \mathbf{a}_{r+1} + \mathbf{p}_{r+1}$ where the vectors $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_{r+1}$ define points of T_{n-r-1}. Hence all these $S^\mu{}_r$ will have dimension r. All the vectors $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_{r+1}$ must approach the null vector as μ approaches infinity otherwise we could construct a subsequence which would contradict the hypothesis. Thus the lemma is proved.

We introduce the following multiplicity convention:

A point P is said to be within the space $(n-1, s)$ exactly k-fold if $P \in (n-k, s)$, P non $\in (n-k-1, s)$, $0 < k < n$, and n-fold if $P = s$.

LEMMA 3. *For $n > 1$, $k > 1$ an arc A of C_n contains points $s_1, s_2, \ldots, s_k$ with $s_1 \leqslant s_2 \leqslant \ldots \leqslant s_k$ and all different from one of its endpoints s_a. P is a space*

point for which $P \neq s_a$ and $P \in (n-1, s_i)$, $1 \leqslant i \leqslant k$. Then the projection of P from s_a will be included within at least $k-1$ spaces $(n-2, s)$ of the projection C_{n-1} of C_n for which $s_1 \leqslant s \leqslant s_k$. Multiple inclusions are to be interpreted in accordance with the multiplicity convention.

Proof. For s on the given arc A of C_n let $Q(s)$ be the intersection of $(n-1, s)$ and the line s_aP; $Q(s)$ is uniquely defined except possibly for $s = s_a$. It is continuous as $(n-1, s)$ is continuous by (2). It cannot cover the full projective line $s_a P$ as $Q(s) \neq s_a$, $s \neq s_a$, for all s in A including the second endpoint. For $i < k$ let $s_i < s_{i+1}$; $Q(s_i) = Q(s_{i+1}) = P$ but $Q(s)$ cannot be equal to P for all s with $s_i < s < s_{i+1}$ by Theorem 2. Hence $Q(s)$ must attain an extremum at a point s'_i for which $s_i < s'_i < s_{i+1}$. Within every curve neighborhood of s'_i two points separated by s'_i must exist for which $Q(s)$ attains the same value. Then by Theorem 1 and the continuity of $Q(s)$, $Q(s) \in (n-2, s'_i)$.

Let m be the number of different values of s_i and let s_j run through each of these different values exactly once. Let n_j be the number of s_i which assume the value s_j. By hypothesis $\sum_j n_j = k$. Let $\overline{P}$ be the projection of P from s_a and C_{n-1} that of C_n. As the space $(n-2, s'_i)$ of C_n projects into the space $(n-2, s'_i)$ of C_{n-1}, $\overline{P} \in (n-1-1, s'_i)$. Similarly, if $P \in (n-n_j, s_j)$ of C_n then $\overline{P} \in (n-1-(n_j-1), s_j)$ of C_{n-1}. Hence $\overline{P}$ is contained in at least $m-1+\sum_j (n_j-1) = k-1$ spaces $(n-2, s)$ of C_{n-1} for which $s_1 \leqslant s \leqslant s_k$. Thus the lemma is proved.

THEOREM 3. *No space point P is within more than n spaces $(n-1, s)$ of C_n.*

Proof. This theorem is the statement that the dual of C_n has property (1). As C_1 is self-dual it is true for C_1. We assume the result for curves C_{n-1} and proceed by induction. If the result is false for a curve C_n then an arc of this curve exists with distinct endpoints s_a, s_b together with $n+1$ points $s_1, s_2, \ldots, s_{n+1}$ with $s_a \leqslant s_1 \leqslant s_2 \leqslant \ldots \leqslant s_{n+1} \leqslant s_b$ so that $P \in (n-1, s_i)$, $1 \leqslant i \leqslant n+1$. Multiple inclusions are interpreted in accordance with the multiplicity convention. P cannot be the point s_a for in this case P would be included in $(n-1, s_a)$ n-fold and by (1) (with the added multiplicity convention) in no other spaces $(n-1, s)$. Let P be included in $(n-1, s_a)$ k-fold, $0 \leqslant k < n$ where $k = 0$ is to be interpreted as P non $\in (n-1, s_a)$. Then P is contained in $n-k+1$ spaces $(n-1, s)$ with $s \neq s_a$. If we project from s_a then the projection $\overline{P}$ of P will, by Lemma 3, be contained in at least $n-k$ spaces $(n-2, s)$ of the projected curve C_{n-1} in addition to being contained in $(n-2, s_a)$ k-fold. In all, $\overline{P}$ is contained in at least n spaces $(n-2, s)$ of C_{n-1} in contradiction to the induction assumption. Hence P can be contained in at most n spaces $(n-1, s)$ and the theorem is proved.

THEOREM 4. *Points $s^\mu_1, s^\mu_2, \ldots, s^\mu_{k+1}$ are defined for $\mu = 0, 1, 2, 3, \ldots,$ and all converge to $\bar{s}$ as μ approaches infinity. Then the intersection S^μ of the*

spaces $(n - 1, s^\nu_1)$, $(n - 1, s^\nu_2), \ldots, (n - 1, s^\nu_{k+1})$, $0 \leqslant k < n$, *converges to* $(n - k - 1, \bar{s})$. *The points of* S^ν *are to be included h-fold within any hyperplane which occurs h times in this set.*

Proof. The theorem is the statement that the dual of (2) is true for $\mathfrak{C}_n$. For $k = 0$ the result is a statement of the continuity of $(n - 1, s)$ which we assume by (2). In particular the result is true for C_1. Therefore let $k > 0$. We assume the result for C_{n-1} and proceed by induction. We select a point P^ν from each S^ν. As the dimension of $S^\nu \geqslant n - k - 1$ the truth of the theorem will result from Lemma 2 if we prove that every convergent subsequence P^{ν} of P^ν has its limit P within $(n - k - 1, \bar{s})$. We may assume $s^\nu_1 \leqslant s^\nu_2 \leqslant \ldots \leqslant s^\nu_{k+1}$. With the help of Theorem 2 we select an arc A containing $\bar{s}$ for one of the endpoints s_a of which $\bar{s} \neq s_a$ and P non $\in (n - 1, s_a)$. If we choose P^{ν} sufficiently close to P, we may assume P^{ν} non $\in (n-1, s_a)$ and also, if $s^{\nu}_1, s^{\nu}_2, \ldots, s^{\nu}_{k+1}$ are sufficiently close to $\bar{s}$, that these points will be within A and different from s_a. Let $\bar{P}$ be the projection of P from s_a, C_{n-1} that of C_n and $\bar{P}^{\nu}$ that of P^{ν}. By Lemma 3, $\bar{P}$ will be contained in k spaces $(n - 2, s)$ of C_{n-1} with $s^{\nu}_1 \leqslant s \leqslant s^{\nu}_{k+1}$. $\bar{P}^{\nu}$ will converge to $\bar{P}$ and, by the induction assumption applied to C_{n-1}, $\bar{P} \in (n - 1 - k, \bar{s})$. Therefore P is contained in the space generated by s_a and $(n - k - 1, \bar{s})$ of C_n. If P non $\in (n-k-1, \bar{s})$. then s_a will be in the space generated by P and $(n - k - 1, \bar{s})$. As s_a may be chosen in infinitely many ways this would contradict the assumption (1). Hence $P \in (n - k - 1, \bar{s})$. The theorem is then completely proved.

I should like to take the opportunity to thank Dr. O. Haupt and Dr. P. Scherk for the improvements they suggested in my original manuscript.

University of British Columbia

THE MATHIEU GROUPS

R. G. STANTON

1. Introduction. An enumeration of known simple groups has been given by Dickson [17]; to this list, he made certain additions in later papers [15], [16]. However, with but five exceptions, all known simple groups fall into infinite families; these five unusual simple groups were discovered by Mathieu [21], [22] and, after occasioning some discussion [20], [23], [27], were relegated to the position, which they still hold, of freakish groups without known relatives. Further interest is attached to these Mathieu groups in virtue of their providing the only known examples (other than the trivial examples of the symmetric and alternating groups) of quadruply and quintuply transitive permutation groups.

Basically, the two important Mathieu groups are the group $\mathfrak{M}_{12}$ of order $m_{12} = 95040$ and the group $\mathfrak{M}_{24}$ of order $m_{24} = 244823040$. The other three Mathieu groups are subgroups of these two; $\mathfrak{M}_{11}$ is a subgroup of index 12 in $\mathfrak{M}_{12}$ whereas $\mathfrak{M}_{23}$ is a subgroup of index 24 in $\mathfrak{M}_{24}$ and $\mathfrak{M}_{22}$ is a subgroup of index 23 in $\mathfrak{M}_{23}$. Since the Mathieu groups are exceptional both in their simplicity and their multiple transitivity, it should be of interest to investigate whether they are the unique simple groups of their orders. Hence we shall consider the two groups $\mathfrak{M}_{12}$ and $\mathfrak{M}_{24}$ and prove the following

MAIN THEOREM. *The only simple group of order m_{12} is the Mathieu group $\mathfrak{M}_{12}$; the only simple group of order m_{24} is the Mathieu group $\mathfrak{M}_{24}$.*

2. Definition of the Mathieu groups. A brief summary of the history of the Mathieu groups is provided by Witt [30]; we shall ignore the older permutation definition and give a more combinatorial one which is also due to Witt. This necessitates the introduction of the concept of a Steiner system [31].

A Steiner system $\mathfrak{S}(l, m, n)$ is defined to be a set of $\binom{n}{l}/\binom{m}{l}$ m-member clubs formed from n individuals who are subject to the proviso that every l persons must meet together in one club and one club only. Clearly one person will occur in $m\rho/n$ clubs, ρ denoting the total number of clubs. Such a Steiner system is identical with the "complete 1-l-m configuration" of tactical arrangement [13], [24], [25].

The Steiner group of $\mathfrak{S}(l, m, n)$ is the group of all those permutations of the n individuals which leaves the set of clubs invariant; such a group will be l-fold transitive. We now define the two fundamental Mathieu groups $\mathfrak{M}_{12}$ and $\mathfrak{M}_{24}$ as the groups of the two Steiner systems $\mathfrak{S}(5, 6, 12)$ and $\mathfrak{S}(5, 8, 24)$.

Received March 6, 1950.

This paper contains the substance of a thesis prepared under the supervision of Professor Richard Brauer and submitted for a Ph.D. degree at the University of Toronto in May 1948.

If we consider all the clubs of $\mathfrak{S}(l, m, n)$ which contain a fixed individual, these clubs will in turn form another Steiner system $\mathfrak{S}(l\text{-}1, m\text{-}1, n\text{-}1)$. The systems $\mathfrak{S}(4, 5, 11)$ and $\mathfrak{S}(4, 7, 23)$, $\mathfrak{S}(3, 6, 22)$ can thus be obtained from the two systems given in the preceding paragraph. Their groups, which we call $\mathfrak{M}_{11}$, $\mathfrak{M}_{23}$, and $\mathfrak{M}_{22}$, are the other three Mathieu groups and occur as subgroups of the two larger groups $\mathfrak{M}_{12}$ and $\mathfrak{M}_{24}$.

The orders of $\mathfrak{M}_{12}$ and $\mathfrak{M}_{24}$ may be obtained from the given definitions; thus $\mathfrak{M}_{12}$ is 5-fold transitive and so has order m_{12} equal to 12.11.10.9.8.k, where k is the order of the subgroup leaving five persons invariant. Since every permutation other than the identity of an l-fold transitive group must alter at least $2l - 2$ symbols, it is necessary that k be unity and so $m_{12} = 95040$. A slightly more intricate combinatorial analysis yields $m_{24} = 24.23.22.21.20.48 = 244823040$.

We may now amplify the statement of our problem. Dickson [17] has shown that there are infinitely many group orders g with the property that there exists two simple groups of order g; the lowest such value of g is 20160. We shall here study simple groups of orders m_{12} and m_{24}; it is first shown that the character tables for simple groups of these two orders are unique and are consequently identical with the character tables already known [18], [19] for $\mathfrak{M}_{12}$ and $\mathfrak{M}_{24}$. Thence it is possible to demonstate the theorem stated in the introduction.

3. Modular characters of groups. Since the assumption of the existence of simple groups of order m_{12} and m_{24} is a rather meagre one with which to start, we must attack the problem by local methods. These are provided by the theory of modular group characters for a fixed prime p. We briefly sketch here some of the more fundamental results which we shall employ from this theory.

Consider a group $\mathfrak{G}$ of order $g = p^a q^b r^c \ldots$ where p, q, and r are distinct primes. Let k denote the number of classes of conjugate elements. If a class contains elements of order prime to p, it is called a p-regular class; otherwise it is termed p-singular. It is well known that if $\mathfrak{G}$ is represented by matrices with coefficients in a field K of characteristic prime to the order g of $\mathfrak{G}$, then the ordinary Frobenius-Schur theory of representation holds and there are k irreducible representations over K [14] with corresponding irreducible characters $\zeta_1, \zeta_2, \ldots, \zeta_k$ [12], [26], [28]. But if K has characteristic p, where p divides g, this theory is no longer valid; the ordinary irreducible representations, when their coefficients are taken in such a modular field K, break up further into modular-irreducible representations and the number of these is equal to the number of p-regular classes of conjugate elements. This splitting actually corresponds to the fact that the group algebra Γ of $\mathfrak{G}$ is not semi-simple when taken as an algebra over the field K. The traces of the modular-irreducible representations are, after an isomorphic mapping upon the roots of unity in the complex field, referred to as the modular-irreducible characters.

This splitting-up, in the field K, of the ordinary irreducible representations into modular-irreducible representations allows us to make a very significant grouping of the ordinary representations. A set of ordinary irreducible representations is said to form a *block for the prime* p if they can be written down in some chain order such that each representation has a modular-irreducible constituent in common with both the preceding and the following representations (for brevity, we also say that the corresponding characters belong to the same block of characters). Such blocks may run the whole gamut of possibilities from blocks made up of a single ordinary representation to blocks consisting of all the ordinary representations. The theory of modular characters and blocks is developed in detail in [1], [2], [3], [6], [7], [8], [10], [11].

We shall denote the various blocks by the notation $B_\mu(p)$ and agree that $B_1(p)$ shall refer to that block containing the unit representation. The *type* of a block is defined as the minimal power of p dividing the degrees of all representations in that block; it may range from 0 to a and is always 0 for $B_1(p)$. At the present state of our knowledge, the most useful blocks are those of type a-1; we refer to these as *standard blocks*.

4. Structure of the blocks. If an ordinary irreducible representation has degree divisible by p^a, then it remains modular-irreducible and forms a block by itself. In particular, if $g = pg'$, where $(p, g') = 1$, we have an especially simple situation which has been extensively studied in [4], [5], [29]; all representations are either individual blocks or else fall into standard blocks of representations whose degrees are all prime to p.

In this particular case where p divides g to the first power only, the group order g may be written in the form

$$g = p\,\frac{p-1}{t}\,v(1+np) \tag{1}$$

where pv is the order of the normalizer of the element P of order p, t is the number of different classes of conjugate elements of $\mathfrak{G}$ appearing in the Sylow subgroup $\{P\}$, and $1 + np$ is the number of Sylow subgroups of order p. The standard blocks $B_\mu(p)$ consist of $(p-1)/t_\mu$ ordinary characters (t_μ being a divisor of $p-1$) and a family of t_μ p-conjugate characters, that is, characters which differ only in a permutation of the p-th roots of unity, the g'-th roots of unity remaining unaltered. In particular, $t_1 = t$.

The normalizer of P can be written as $\{P\} \times \mathfrak{V}$ where $\mathfrak{V}$ is a group of order v, this order usually being small. If the characters of $\mathfrak{V}$ are known, then we can find all the characters of $\mathfrak{G}$ itself insofar as they lie in standard blocks [4]. In particular, all the characters of $B_1(p)$ have degrees which are congruent to ± 1 modulo p except for one exceptional family of t p-conjugate characters whose members have degrees congruent to $\pm (p-1)/t$ modulo p. The other blocks $B_\mu(p)$ contain characters whose degrees are congruent to $\pm a_\mu$ modulo p and a family of t_μ p-conjugate characters whose members have degrees con-

gruent to $\pm a_\mu(p-1)/t_\mu$ modulo p. In any one of these blocks we may consider the characters to be of two kinds; those with degrees congruent to a_μ (including the exceptional family if its t_μ members have degrees congruent to $-a_\mu(p-1)/t_\mu$) are said to be of positive type, and those with degrees congruent to $-a_\mu$ (including the exceptional family if its members have degrees congruent to $a_\mu(p-1)/t_\mu$) are said to be of negative type. If we set the sum of all characters of positive type equal to the sum of all characters of negative type, we obtain a character relation which is valid for the p-regular classes of elements. This relation will be our most powerful tool.

5. Block relationships for two primes. In this section we shall briefly enumerate some of the most useful theorems on blocks and block-intersections. We use, as before, p and q to denote two primes which divide the group order g.

LEMMA 1. *If a relation*

$$\sum_{\mu=1}^{k} a_\mu \zeta_\mu(S) = 0$$

holds for all p-singular elements S of $\mathfrak{G}$, a_μ being independent of S, then the relation still holds if the summation is performed only over characters of some fixed block B, that is,

$$\sum_{\zeta_\mu \in B} a_\mu \zeta_\mu(S) = 0.$$

Proof. Determine numbers b_G such that $a_\mu = \sum_{G \in \mathfrak{G}} b_G \zeta_\mu(G)$. The orthogonality relations for ordinary characters show that $b_G = 0$ for p-singular elements G. Hence $a_\mu = \sum_R b_R \zeta_\mu(R)$ where R ranges over the p-regular elements of G. Then

$$\sum_{\zeta_\mu \in B} a_\mu \zeta_\mu(S) = \sum_R b_R \sum_{\zeta_\mu \in B} \zeta_\mu(R)\zeta_\mu(S) = 0.$$

(Cf. [10], Theorem 8).

LEMMA 2. *If $\mathfrak{G}$ contains no elements of order pq and if*

$$\sum_{\mu=1}^{k} a_\mu \zeta_\mu(G) = 0$$

for all p-regular elements G, then

$$\sum_{\zeta_\mu \in B} a_\mu \zeta_\mu(H) = 0$$

for all q-singular elements H, the summation being performed over the characters of a fixed q-block B. Furthermore, if E is the identity in $\mathfrak{G}$, then

$$\sum_{\zeta_\mu \in B} a_\mu \zeta_\mu(E) \equiv 0 \bmod q^b.$$

Proof. (The a_μ denote algebraic integers). Every q-singular H is p-regular and so the hypothesis gives $\sum_{\mu=1}^{k} a_\mu \zeta_\mu(H) = 0$. Apply Lemma 1 for the prime q

and we have $\sum_{\zeta_\mu \in B} a_\mu \zeta_\mu(H) = 0$. In particular, the expression $S(X) = \sum_{\zeta_\mu \in B} a_\mu \zeta_\mu(H)$ vanishes for all elements, except the identity, of a Sylow q-group $\mathfrak{Q}$. Express $S(X)$ as a linear combination of the irreducible characters of $\mathfrak{Q}$, for X in $\mathfrak{Q}$. The coefficient of the principal character of $\mathfrak{Q}$ is $q^{-b}S(E)$; as this number is an algebraic integer, $S(E) \equiv 0 \bmod q^b$, that is, $\sum_{\zeta_\mu \in B} a_\mu \zeta_\mu(E) \equiv 0 \bmod q^b$.

Lemma 3. *If a character ζ belongs to the first p-block, so do all its algebraically conjugate characters.*

Proof. The necessary and sufficient condition for ζ to belong to the first p-block is

$$\frac{g}{n(G)} \frac{\zeta(G)}{\zeta(E)} \equiv \frac{g}{n(G)} \bmod \mathfrak{p}$$

for all G in $\mathfrak{G}$, where $\mathfrak{p}$ is a prime ideal dividing p and $n(G)$ is the order of the normalizer of G. An algebraically conjugate character ζ' can be obtained by replacing G by G^α where $(\alpha, g) = 1$. Then $n(G^\alpha) = n(G)$ and, using this condition, with the relation already given, for the element G^α, we have

$$\frac{g}{n(G)} \frac{\zeta'(G)}{\zeta'(E)} \equiv \frac{g}{n(G)} \bmod \mathfrak{p}.$$

This shows ζ' is in the first p-block.

In the following lemmas, we consider groups of orders divisible by a prime p to the first power only, that is, a decomposition of g exists in the form (1).

Lemma 4. *If $v > 1$, the degrees z of all characters other than the 1-character belonging to $1-1$ representations in the first p-block satisfy the inequality $z \geqslant 1 + 2p$.* (In this lemma, we assume $\mathfrak{G} = \mathfrak{G}'$).

Proof. We have g written in the form (1) with $v > 1$. Let ζ be a character of the first p-block; then $\zeta(V) \equiv \zeta(PV) = \zeta(P) \equiv z \bmod \mathfrak{p}$. If we consider ζ as a character of $\mathfrak{B}$, we have $\zeta(\mathfrak{B}) = a(1) + \sum c_\nu\theta_\nu$ where the θ_ν are irreducible characters of $\mathfrak{B}$ with $\theta_\nu \neq (1)$, $a \geqslant 0$, $c \geqslant 0$, a and c rational integers. Then $\zeta(\mathfrak{B}) \equiv z \bmod \mathfrak{p}$.

Now the order v is prime to p and so $a \equiv z \bmod p$, $c_\nu \equiv 0 \bmod p$. Then $\zeta(\mathfrak{B}) = a(1) + p\sum b_\nu\theta_\nu$ and $z = a + p\sum b_\nu = a + pb$. Since $z \not\equiv 0 \bmod p$, $a > 0$; also, if $b = 0$, all $b_\nu = 0$. In this case, $\zeta(\mathfrak{B}) = a(1)$ and $\mathfrak{B}$ is represented by the identity; thus $b \geqslant 1$.

If $b = 1$, then the representation corresponding to ζ is composed of the identity of order a along with p repetitions of a linear representation $\mathfrak{F}$, that is,

$$Z(\mathfrak{B}) = \begin{pmatrix} I_a & \\ & p \times \mathfrak{F} \end{pmatrix}.$$

Then $\det Z(\mathfrak{B}) = \{\det \mathfrak{F}(\mathfrak{B})\}^p$. At this stage, we make the assumption that $\mathfrak{G}$ is identical with its derived group $\mathfrak{G}'$. Then $\det Z$ is a linear character of

$\mathfrak{G}$ and so is unity; hence det $\mathfrak{F}(\mathfrak{B}) = 1$. But $\mathfrak{F}(\mathfrak{B})$ is linear and so $\mathfrak{F}(\mathfrak{B}) = 1$; this is impossible. Thus $b \geqslant 2$ and so

$$z = a + pb \geqslant 2p + 1$$

for all characters of $1 - 1$ representations in the first p-block (when $\mathfrak{G} = \mathfrak{G}'$).

LEMMA 5. *Suppose that p occurs in g to the first power only, as in* (1); *Let the decomposition* (1) *for a second prime p' be $g = p'\left(\frac{p'-1}{t'}\right)v'(1 + n'p')$. Let the group $\mathfrak{B}'$ of order v' intersect $\mathfrak{B}$ in a group $\mathfrak{W}$ of order w. Then, if $w \neq 1$, every representation in $B_1(p) \cap B_1(p')$ has a degree z satisfying the inequality $z \geqslant 1 + 2pp'$.*

Proof. The proof parallels closely that of Lemma 4. Let W be an element of $\mathfrak{W} = \mathfrak{B} \cap \mathfrak{B}'$. Then, by the argument of Lemma 4, we have

$$\zeta(\mathfrak{W}) = a(1) + p\omega(\mathfrak{W})$$

where ω is a character, reducible or irreducible, of $\mathfrak{W}$. In a similar manner,

$$\zeta(\mathfrak{W}) = a'(1) + p'\omega'(\mathfrak{W}).$$

These two expressions for $\zeta(\mathfrak{W})$ may then be equated. Now let ω_0 be a character of $\mathfrak{W}$, other than the principal character, which appears in ω; its multiplicity must be divisible by pp'. Then

$$\zeta(\mathfrak{W}) = \mathfrak{a}(1) + pp'\omega_1(\mathfrak{W}).$$

Taking degrees, and using the same sort of determinantal argument as in Lemma 4, we obtain the inequality

$$z \geqslant \mathfrak{a} + 2pp'$$

where $\mathfrak{a} > 0$, $\mathfrak{a} \equiv a \bmod p$, $\mathfrak{a} \equiv a' \bmod p'$.

LEMMA 6. *As in Lemma 5, assume that p and p' are distinct primes which divide g to the first power only and that there are no elements of order pp'. Let a_{ij} be the number of characters in $B_1(p) \cap B_1(p')$ which are of type i for p and type j for p', the indices i and j being zero or unity according as the character type, defined at the end of* §4, *is positive or negative. Then*

$$a_{00} + a_{11} = a_{01} + a_{10}. \tag{2}$$

Proof. Let a character ζ be in common to the first p-block and the first p'-block; three cases may arise. First ζ may be non-exceptional for both p and p'; in this case the degree z of ζ is congruent to ± 1 for both p and p'. Secondly, ζ may be exceptional for p', that is, $z \equiv \pm 1 \bmod p$, $z \equiv \mp \frac{p'-1}{t'} \bmod p'$. Thirdly, ζ may be exceptional for p, that is, $z \equiv \pm 1 \bmod p'$, $z \equiv \mp \frac{p-1}{t} \bmod p$. ζ can not be exceptional for both p and p'.

Now consider the degree relation for the first block $B_1(p)$ and take the degrees modulo p'. The degrees which are of positive type for p' will then contribute 1 to the congruence; those which are of negative type will contribute -1. In this way we obtain the congruence

$$a_{00} - a_{01} \equiv a_{10} - a_{11} \mod p'.$$

If $t' > 1$, the sum $a_{00} + a_{01} + a_{10} + a_{11} \leqslant \dfrac{p'-1}{t'}$ and so the above congruence must be an equality. If $t' = 1$, then $a_{00} + a_{01} + a_{10} + a_{11} \leqslant p'$ and the congruence must again be an equality. Thus we have the block-intersection theorem

$$a_{00} + a_{11} = a_{01} + a_{10}.$$

LEMMA 7. *Let p, p', w have the same significance as in Lemma* 5. *Assume now that p' divides $\dfrac{p-1}{t}$ and that $w = 1$. Then $v \equiv 1 \bmod p'$ and the number of elements of $\mathfrak{B}$ of a fixed order other than* 1 *is also congruent to* 1 mod p'.

Proof. Let $\mathfrak{M}$ denote the normalizer of $\{P\}$ and let P' denote an element of order p'. Since $\mathfrak{B}$ is normal in $\mathfrak{M}$, P' must transform $\mathfrak{B}$ into itself. Also $w = 1$; hence 1 is the only invariant element and this implies that v is congruent to 1 mod p'.

Consider now a class of conjugate elements, other than the identity, in $\mathfrak{B}$; P' will transform this class into another class. If this second class were not distinct from the first, then the number of elements in it would have to be congruent to zero mod p'. This is not possible, under the assumptions of the lemma, since p' can not divide v. Thus P' carries a class of $\mathfrak{B}$ into a distinct class and this completes the Lemma.

COROLLARY; *Under the assumptions of Lemma* 7, *but without insisting that $w = 1$, we have $w \equiv v$* mod p'.

In concluding this section, it might be well to emphasize that, while we are here applying the powerful local methods of modular theory to the Mathieu groups, the same general approach could be used on any members of the rather large class of groups which contain a prime (or preferably several primes) to the first power only.

6. The blocks in the Mathieu groups. It is not possible to give here[1] the considerable numerical calculations necessary to find the degrees of the characters of simple groups of orders m_{12} and m_{24}. We shall content ourselves with indicating briefly the process for the case of the prime 23 in a simple group of order m_{24}. We know that there is a decomposition (1) of m_{24} and, by considering the factors of m_{24} mod 23, the number $1 + np$ of Sylow 23-

[1]The details of numerical calculation are available in the author's thesis in the University of Toronto library.

groups is found to lie in the set 24, 70, 231, 576, 990, 1680, 3520, 5544, 13824, 19712, 23760, 40320, 84480, 133056, 967680. Also, the number t can, for a simple group, be only 1, 2, or 11. The possibility $t = 11$ can be excluded almost immediately and this in turn eliminates some of the numbers in the list of possible Sylow groups. From there on, numerical work with Lemmas 4, 5, 6, and 7 is necessary. The block-intersection theorem that there must be a character other than the 1-character in $B_1(23) \cap B_1(11)$ finally allows us to eliminate all cases except $t = 2$, $1 + np = 967680$. Thus the decomposition (1) for the prime 23 is

$$m_{24} = 23.11.1.967680.$$

The decomposition of m_{24} in the form (1) for the other primes 11, 7, and 5 which appear in the group order to the first power requires an even more extended numerical sieving of possible degrees in block-intersections such as $B_1(23) \cap B_1(5)$, etc. When completed, the decompositions are

$$\begin{aligned} m_{24} &= 11.10.1.1225664 && \text{for } 11,\\ m_{24} &= 5.4.12(23.11.7.576) && \text{for } 5,\\ m_{24} &= 7.3.6(23.11.5.1536) && \text{for } 7. \end{aligned}$$

The groups $\mathfrak{B}$ of orders 12 and 6 which appear associated with the primes 5 and 7 are just the alternating group on 4 symbols and the symmetric group on three symbols. From the groups $\mathfrak{B}$ we see that there exists a 23-block, an 11-block, 3 5-blocks, and 3 7-blocks. Such a large number of blocks makes for smooth working of the block-intersection theorem and we give the results of its application in the form of block relations (asterisks denote the families of p-conjugate characters; also, we use the convention that the number a shall mean "the character whose degree is a"):

$$\begin{aligned}
&B_1(23) && 1 + 231' + 231'' + 990' + 990'' + 3520 + 5544\\
& && \qquad = 770^* + 45' + 45'' + 252 + 10395,\\
&B_1(11) && 1 + 45' + 45'' + 1035 + 1035' + 1035'' + 23 + 3312\\
& && \qquad = 252 + 483 + 5796,\\
&B_1(7) && 1 + 2024 = 990^* + 1035,\\
&B_2(7) && 3312 + 253 = 3520 + 45^*,\\
&B_3(7) && 23 + 2277 = 1035^* + 1265,\\
&B_1(5) && 1 + 5796 + 1771 = 2024 + 5544,\\
&B_2(5) && 252 + 231^* = 483,\\
&B_3(5) && 23 + 253 + 5313 = 3312 + 2277.
\end{aligned}$$

During the course of the block determination, it also appeared that there were four standard 3-blocks, which were very helpful in finding the character relations, namely:

$$\begin{aligned}
&B_2(3) && 5796 = 252 + 5544,\\
&B_3(3) && 990' + 45' = 1035',\\
&B_4(3) && 990'' + 45'' = 1035'',\\
&B_4(3) && 2277 + 1035 = 3312.
\end{aligned}$$

In the twelve blocks which have just been given there occur 26 degrees; finding the sum of their squares, we check that it is equal to m_{24} and so we have found all the characters. The complete character table can then be constructed by using these block relations, together with the results of [4] concerning the expression of the characters of $\mathfrak{G}$ in terms of those of $\mathfrak{B}$. It turns out that this table can be formed in a unique way, that is, we have

THEOREM 1. *The character table for any simple group of order m_{24} is unique and hence is identical with that for $\mathfrak{M}_{24}$.*

The analogous result for $\mathfrak{M}_{12}$, namely, that the character table for a simple group of order m_{21} is unique, has already been given in [5]. However, we should here give the decomposition (1), which is:

$$m_{12} = 11.5.1.1728 \qquad \text{for } 11,$$
$$m_{12} = 5.4.2.2376 \qquad \text{for } 5.$$

Thus there is an 11-block and 2 5-blocks; in order for them to fit together, we we find that the block relations must be:

$B_1(11)$ $\quad 1 + 45 + 144 = 16^* + 54 + 120,$
$B_1(5)$ $\quad 1 + 66 + 176 = 99 + 144,$
$B_2(5)$ $\quad 16' + 16'' + 11' + 11'' = 54.$

There is also a standard 3-block and a 2-block type 4; these are given by:

$B_3(3)$ $\quad 45 + 99 = 144,$
$B_1(2)$ $\quad 16' + 16'' + 144 = 176.$

The fifteen degrees which occur in these block relations suffice to fill up the group order 95040 and, as in the case of m_{24}, the character table can be uniquely constructed from the block relations.

7. Uniqueness of the Mathieu groups. It is a well-known fact that two distinct groups of a given order g may possess the same character table; we now seek to show that this can not be the case for m_{12} or m_{24}. Suppose that we consider the character table for $\mathfrak{M}_{12}$ (for a reproduction of this table, cf. [19]). Let the corresponding group be represented as a group of linear substitutions in 11 variables x_i; by a rather lengthy discussion of the canonical matric form of the elements of order 11 and order 5, one can show that the invariance group of the variable x_1 is a group of order 7920. When this is done, the proof runs smoothly; the group under consideration must have a subgroup of index 12 and hence a permutation representation of degree 12. Split this permutation representation into irreducible constituents; the only possible splitting is a splitting into the unit representation and a representation of degree 11. This is, however, a necessary and sufficient condition for the double transitivity of the group. An exactly similar discussion of a simple group of order m_{24} can be carried out using the representation of degree 23; the invariance group of x_1 will have index 24 in this case. Hence we obtain

THEOREM II. *A simple group of order m_{12} is doubly transitive on 12 symbols; a simple group of order m_{24} is doubly transitive on 24 symbols.*

By consulting the tables of primitive groups, we could immediately identify the group on 12 symbols as $\mathfrak{M}_{12}$; however, these tables do not extend as far as degree 24 and so it is better to proceed by writing down permutation representations for the group elements of these doubly transitive groups. When these are obtained, they turn out to be identical with the known permutation representations of $\mathfrak{M}_{12}$ and $\mathfrak{M}_{24}$ [13], [20], [27]. This result, combined with Theorems I and II, yields the main theorem, as given at the end of §1.

REFERENCES

[1] R. Brauer, *On the Cartan Invariants of Groups of Finite Order*, Ann. of Math., vol. 42 (1941), 53-61.

[2] ———, *On the Connection Between the Ordinary and the Modular Characters of Groups of Finite Order*, Ann. of Math., vol. 42 (1941), 926-935.

[3] ———, *Investigations on Group Characters*, Ann. of Math., vol. 42 (1941), 936-958.

[4] ———, *On Groups whose Order Contains a Prime Number to the First Power*, Am. J. of Math., vol. 64 (1942), 401-440.

[5] ———, *On Permutation Groups of Prime Degree and Related Classes of Groups*, Ann. of Math., vol. 44 (1943), 57-79.

[6] ———, *On the Arithmetic in a Group Ring*, Proc. Nat. Acad. Sciences, vol. 30 (1944), 109-114.

[7] ———, *On Blocks of Characters of Groups of Finite Order*, Proc. Nat. Acad. Sciences, vol. 32 (1946), 182-186 and 215-219.

[8] ———, *On Modular and p-adic Representations of Algebras*, Proc. Nat. Acad. Sciences, vol. 25 (1939), 252-258.

[9] ———, *On the Representation of a Group of Order g in the Field of the g-th Roots of Unity*, Am. J. of Math., vol. 67 (1945), 461-471.

[10] R. Brauer and C. Nesbitt, *On the Modular Representations of Groups of Finite Order*, Univ. of Toronto Studies, No. 4 (1937).

[11] ———, *On the Modular Characters of Groups*, Ann. of Math., vol. 42 (1941), 556-590.

[12] W. Burnside, *The Theory of Groups of Finite Order*, Cambridge (1911).

[13] R. Carmichael, *An Introduction to the Theory of Groups of Finite Order*, Boston (1937).

[14] L. Dickson, *On the Group Defined for any Given Field by the Multiplication Table of any Finite Group*, T.A.M.S., vol. 3 (1902), 285-301.

[15] ———, *Theory of Linear Groups in an Arbitrary Field*, T.A.M.S., vol. 2 (1901), 363-394.

[16] ———, *A new System of Simple Groups*, Math. Ann., vol. 60 (1905), 137-150.

[17] ———, *Linear Groups*, Leipzig (1901).

[18] G. Frobenius, *Über die Charactere der Symmetrischen Gruppe*, Sitz. Preuss. Akad. Wissen. (1900), 516-534.

[19] ———, *Über die Charactere der Mehrfach transitiven Gruppen*, Sitz. Preuss. Akad. Wissen. (1904), 558-571.

[20] C. Jordan, *Traité des Substitutions*, Paris (1870).

[21] E. Mathieu, *Mémoire sur l'étude des fonctions de plusieurs quantités*, Jour. de Math., 2me Série, vol. 6 (1861), 241-323.

[22] ———, *Sur la fonction cinq fois transitive de 24 quantités*, Jour. de Math., 2me Série, vol. 18 (1873), 25-46.

[23] G. Miller, *On the Supposed Five-fold Transitive Function of 24 Elements*, Mess. of Math., vol. 27 (1898), 187-190.
[24] E. Moore, *Tactical Memoranda*, Am. J. of Math., vol. 18 (1896), 268-275.
[25] E. Netto, *Lehrbuch der Kombinatorik*, Leipzig (1927).
[26] I. Schur, *Neue Begründung der Theorie der Gruppencharactere*, Sitz. Preuss. Akad. Wissen. (1905), 406-432.
[27] J. de Séguier, *Theorie des Groupes Finis*, Paris (1904).
[28] A. Speiser, *Die Theorie der Gruppen*, New York (1945).
[29] H. Tuan, *On Groups whose Orders Contain a Prime to the First Power*, Ann. of Math., vol. 45 (1944), 110-140.
[30] E. Witt, *Die 5-fach Transitiven Gruppen von Mathieu*, Abhand. Math. Sem. Hamb., Band 12 (1938), 256-264.
[31] ———, *Über Steinersche Systeme*, Abhand. Math. Sem. Hamb., Band 12 (1938), 265-275.

The University of Toronto

METRIZATION OF TOPOLOGICAL SPACES

R. H. BING

A single valued function $D(x, y)$ is a *metric* for a topological space provided that for points x, y, z of the space:

1. $D(x, y) \geqslant 0$, the equality holding if and only if $x = y$,
2. $D(x, y) = D(y, x)$ (*symmetry*),
3. $D(x, y) + D(y, z) \geqslant D(x, z)$ (*triangle inequality*),
4. x belongs to the closure of the set M if and only if $D(x, m)$ (m element of M) is not bounded from 0 (preserves limit points).

A function $D(x, y)$ is a metric for a point set R of a topological space S if it is a metric for R when R is considered as a subspace of S. A topological space or point set that can be assigned a metric is called metrizable.

If a topological space has a metric, this metric may be useful in studying the space. Determining which topological spaces can be assigned metrics leads to interesting and important problems. For example, see [3].

A regular[1] topological space is metrizable if it has a countable basis[2] [7 and 8]. However, it is not necessary that a space be separable[3] in order to be metrizable. Theorem 3 gives a necessary and sufficient condition that a space be metrizable by using a condition more general than perfect separability.

Alexandroff and Urysohn showed [1] that a necessary and sufficient condition that a topological space be metrizable is that there exist a sequence of open coverings $G_1, G_2, \ldots$ such that (a) G_{i+1} is a refinement[4] of G_i, (b) the sum of each pair of intersecting elements of G_{i+1} is a subset of an element of G_i, and (c) for each point p and each open set D containing p there is an integer n such that every element of G_n containing p is a subset of D. We call a sequence of open coverings satisfying condition (c) a development. A developable space is a topological space that has a development. In section 2 we study conditions under which developable spaces can be assigned metrics.

The results of this paper hold in a topological space as defined by Whyburn in [10] or in a Hausdorff space.

Received November 26, 1949.

[1]A topological space is regular if for each open set D and each point p in D there is an open set containing p whose closure lies in D.

[2]A basis for a topological space is a collection G of open sets such that each open set is the sum of a subcollection of G. In [7 and 8] a space with a countable basis was said to satisfy the second axiom of countability. More recently such spaces have been called perfectly separable.

[3]A space is separable if it has a countable dense subset.

[4]The collection G is a refinement of the collection H if each element of G is a subset of an element of H.

1. Screenable spaces. We shall use the following definitions:

Discrete. A collection of point sets is *discrete* if the closures of these point sets are mutually exclusive and any subcollection of these closures has a closed sum.

Screenable. A space is *screenable* if for each open covering H of the space, there is a sequence $H_1, H_2, \ldots$ such that H_i is a collection of mutually exclusive domains and $\sum H_i$ is a covering of the space which is a refinement of H. A space is *strongly screenable* if there exist such H_i's which are discrete collections.

Perfectly screenable. A space is *perfectly screenable* if there exists a sequence $G_1, G_2, \ldots$ such that G_i is a discrete collection of domains and for each domain D and each point p in D there is an integer $n(p, D)$ such that $G_{n(p,D)}$ contains a domain which lies in D and contains p.

Collectionwise normal. A space is *collectionwise normal* if for each discrete collection X of point sets, there is a collection Y of mutually exclusive domains covering X^* such that no element of Y intersects two elements of X. We use X^* to denote the sum of the elements in X.

The following result follows from the definitions of perfectly screenable and strongly screenable.

THEOREM 1. *A perfectly screenable space is strongly screenable.*

The following example shows that a developable space may not be screenable.

EXAMPLE A. *A locally connected*[5] *separable Moore space*[6] *S such that no space homeomorphic with the closure of any open set in S is either normal or screenable.* The points of S are the points of the plane and the open sets of S are given in terms of a development $G_1, G_2, \ldots$ which is described as follows. Let $L_1, L_2, \ldots$ be a sequence of horizontal lines whose sum is dense in the plane. Either of the following types of sets is an element of G_i: (a) the interior of a circle with diameter less than $1/i$ which does not intersect $L_1 + L_2 + \ldots + L_i$, (b) $p + I_1 + I_2$ where p is a point of some L_j and I_1, I_2 are interiors of circles of diameter less than $1/2i$ which are tangent to L_j at p on opposite sides of L_j and such that $I_1 + I_2$ does not intersect $L_1 + L_2 + \ldots + L_i$.

That S is locally connected follows from the fact that vertical lines are connected and horizontal lines other than $L_1, L_2, \ldots$ are connected. The elements of each G_i are connected. Since the plane is separable and any set dense in the plane is dense in S, S is separable. The sequence $G_1, G_2, \ldots$ satisfies the conditions of Axiom 1 of [5], so S is a Moore space.

Let S' be a space homeomorphic with the closure of an open set E in S and K be an interval in E that is a subset of $L_1 + L_2 + \ldots$. If K_1 and K_2 are

[5] A topological space is locally connected if it has a basis such that the elements of the basis are connected. A set is connected if it cannot be expressed as the sum of two non-null sets such that neither contains a point of the closure of the other.

[6] A space satisfying the first three parts of Axiom 1 of [5] is called a Moore space. It is a regular developable space.

subsets of S' corresponding to the points of K with rational and irrational abscissas respectively, K_1 and K_2 are two mutually exclusive closed point sets. That S' is not normal follows from the fact that there do not exist two mutually exclusive domains containing K_1 and K_2 respectively.

If H is an open covering of S' such that no two points of $K_1 + K_2$ belong to the same element of H, any open covering of S' that refines H contains uncountably many elements. Since S' is separable, it does not contain an uncountable collection of mutually exclusive domains. Hence, S' is not screenable because there is not a sequence $H_1, H_2, \ldots$ such that H_1 is a collection of mutually exclusive domains and $\sum H_i$ is a covering of S' that refines H.

The proof of the following theorem may be compared with one given by Tychonoff [7] to show that any regular perfectly separable topological space is normal.

THEOREM 2. *A regular strongly screenable space is collectionwise normal.*

Proof. Suppose $\{A_\alpha\}$ is a discrete collection of closed sets. Let K be a collection of open sets covering the space such that the closure of no element of K intersects two elements of $\{A_\alpha\}$. Since the space is strongly screenable, there is a sequence $H_1, H_2, \ldots$ such that H_i is a discrete collection of domains and $\sum H_i$ is a covering of the space which is a refinement of K.

Let $U_{i\beta}$ be the sum of the elements H_i that intersect A_β and $V_{i\beta}$ be the sum of those intersecting elements of $\{A_\alpha\}$ other than A_β. If $D_\beta = U_{1\beta} + (U_{2\beta} - \overline{V}_{1\beta}) + (U_{3\beta} - [\overline{V}_{1\beta} + \overline{V}_{2\beta}]) + \ldots + (U_{i\beta} - \sum_{j=1}^{i-1} \overline{V}_{j\beta}) + \ldots$, then $\{D_\alpha\}$ is a collection of mutually exclusive domains covering $\sum A_\alpha$ such that no element of $\{D_\alpha\}$ intersects two elements of $\{A_\alpha\}$.

It cannot be concluded that a strongly screenable space is normal for a perfectly separable space may not even be regular. Example B shows us that we cannot conclude that a regular screenable space is normal.

EXAMPLE B. *A screenable, point-wise paracompact,*[7] *nonparacompact, nonnormal Moore space with an open covering H such that the star*[8] *of each point with respect to H is metrizable.*

Points are of three types: (a) elements of a countable sequence of points $p_1, p_2, \ldots$; (b) elements of the collection of all continuous functions $f_\alpha(x)$ $(0 < x < 1)$ such that $\frac{1}{2} < f_\alpha(x) < 1$; and (c) ordered triples (p_i, t, f_α) where p_i is a point of type (a), f_α is one of type (b), and t is a positive number less than one.

The open sets of this space are defined by a development $G_1, G_2, \ldots$ which is described as follows. The elements of G_n are of three sorts: p_j $(j = 1, 2, \ldots)$

[7]A topological space is point-wise paracompact if for each open covering H there is an open covering H' such that H' refines H and no point lies in infinitely many elements of H'. It is paracompact if for each open covering H there are open coverings H' and H'' such that H' refines H and no element of H'' intersects infinitely many elements of H'.

[8]The star of a point set A with respect to an open covering H is the sum of the element of of H that intersect A.

plus all points (p_j, t, f_α) $(t < 1/n)$; (p_j, t_0, f_α) plus all points (p_j, t, f_α) $(|t - t_0| < 1/n)$; an element f_α of type (b) plus all points (p_j, t, f_α) where $1 - t < 1/n$ if $j \leqslant n$ and $1 - t < [1 + nf_\alpha(1/j)]/(n + 1)$ if $j > n$.

It is convenient to think of the space as being the collection of points of type (a) plus the collection of points of type (b) plus open unit intervals joining points of type (a) to points of type (b). An element of G_n is either (i) an element of type (a) plus all points that can be joined to it by intervals of lengths less than $1/n$, or (ii) a point of an open interval plus the collection of all points of the open interval that are nearer than $1/n$ to the point or (iii) a point f of type (c) plus all points that can be joined to it by an interval of length less than $x(j)$ where p is on the interval from f_α to p_j and $x(j) = 1/n$ if $j \leqslant n$ and $x(j) = = [1 + nf_\alpha(1/j)]/(n + 1)$ if $j > n$.

The above space is screenable because for any open covering K of it, there is an open covering $K_1 + K_2$ of it which is a refinement of K and such that K_1 is a collection of mutually exclusive domains covering all points of type (a) and a dense set of those of type (c), while K_2 is another such collection covering all points of type (c) not covered by K_1. The space is point-wise paracompact because for each open covering K of it there is an open covering $K_1 + K_2$ such that no point is covered by more than two elements of $K_1 + K_2$.

The space is not normal because for each domain D containing the collection of all points of type (a), there is a point of type (b) which is a limit point of D. To find such a point, let n_i be an integer such that an element of G_{n_i} contains p_i and lies in D. If f_β is a point of type (b) which satisfies $f_\beta(1/i) > 1 - 1/n_i$ $(i = 1, 2, \ldots)$, it is a point of the closure of D. Since the collection of points of type (a) and the collection of points of type (b) are closed sets, and there do not exist mutually exclusive open sets containing these two collections, the space is not normal. Since it is not normal, it is not paracompact.

To show that there is an open covering H such that the star of any point with respect to H is metrizable, let H be any G_i and M be the star of some point with respect to H. There is a connected open set M' that contains M such that M' contains only one point p_j of type (a) and only one point f_α of type (b). If R is the component of $M' - p_j$ containing f_α, $R + p_j$ has a metric $D_1(x, y)$ because $R + p_j$ is homeomorphic with a regular, perfectly separable space. Furthermore, $M' - R$ has a metric $D_2(x, y)$ described as follows: $D_2[p_j, (p_j, t, f_\alpha)] = t$; $D_2[(p_j, t_1, f_\alpha), (p_j, t_2, f_\beta)]$ is $|t_1 - t_2|$ or $t_1 + t_2$ according as f_α is or is not f_β. There is a metric $D(x, y)$ for M' where $D(x, y) = D_1(x, y)$ if x and y are points of R, $D(x, y) = D_2(x, y)$ if x and y are points of $M' - R$, and $D(x, y) = D_1(x, p_j) + D_2(p_j, y)$ if x and y belong to R and $M' - R$ respectively.

The following theorem may be compared with the result of Urysohn [8] which states that a normal perfectly separable topological space is metrizable.

THEOREM 3. *A necessary and sufficient condition that a regular topological space be metrizable is that it be perfectly screenable.*

Proof of sufficiency. Suppose $H_1, H_2, \ldots$ is a sequence of discrete collections of domains such that for each domain D and each point p in D there is an integer $n(p, D)$ such that an element of $H_{n(p,D)}$ lies in D and contains p.

Let K_{ij} be the sum of the elements of H_j whose closures lie in an element of H_i. Since the space is normal, there is a continuous transformation F_{ij} of space into the real numbers between 0 and 1 such that the image of K_{ij} under F_{ij} is 1 and the image of the complement of H^*_i is 0 [9]. For points x, y of the space we define the distance between them to be

$$D(x, y) = \sum\sum \frac{|F_{ij}(x) + R_{ij}(x, y)F_{ij}(y)|}{2^{i+j}}$$

where $R_{ij}(x, y)$ is -1 or 1 according as y does or does not belong to an element of H_i that contains x. It may be found that $D(x, y)$ satisfies the conditions for a metric.

Proof of necessity. A. H. Stone has shown [6] that for each metric space and each positive number ϵ, there is a sequence $R_{\epsilon 1}, R_{\epsilon 2}, \ldots$ such that $\sum R_{\epsilon i}$ covers the space and $R_{\epsilon i}$ is a discrete collection of closed sets each of diameter less than ϵ. A proof of this is also found in Theorem 9 of the present paper. For each element r of $R_{\epsilon i}$ let D_r be a domain covering r such that D_r is of diameter less than ϵ and each point of D_r is more than twice as close to r as to any other element of $R_{\epsilon i}$. If $H_{\epsilon i}$ denotes the collection of all such open sets D_r, $\{H_{\epsilon i}; \epsilon = 1, \frac{1}{2}, \frac{1}{3}, \ldots \text{ and } i = 1, 2, \ldots\}$ becomes on suitable ordering a countable sequence of discrete collections of open sets which insures that a metric space is perfectly screenable.

In the following modification of Theorem 3 we dispense with the supposition that the elements of H_i are mutually exclusive. E. E. Floyd suggested that such a modification might be possible.

THEOREM 4. *A regular topological space S is metrizable if and only if there is a sequence $G_1, G_2, \ldots$ such that*

(a) *G_i is a collection of open subsets of S such that the sum of the closures of any subcollection of G_i is closed and*

(b) *if p is a point and D is an open set containing p there is an integer $n(p, D)$ such that an element of $G_{n(p,D)}$ contains p and each element of $G_{n(p,D)}$ containing p lies in D.*

Proof. Since a metric space is perfectly screenable, it contains a sequence $G_1, G_2, \ldots$ satisfying the conditions of the theorem. We complete the proof of Theorem 4 by showing that any regular topological space admitting such a sequence is perfectly screenable.

First, we show that any open subset D of S is strongly screenable. Let $H = (h_1, h_2, \ldots, h_a, \ldots)$ be a well ordered collection of open sets whose sum is D. Let V_{ai} be the sum of the elements of G_i whose closures lie in h_a. If U_{aij} denotes the sum of the elements of G_j whose closures lie in V_{ai} but do not intersect $\sum_{\beta < a} \overline{V}_{\beta i}$ then $W_{ij} = \{U_{\gamma ij}; \gamma = 1, 2, \ldots, a, \ldots\}$ is a discrete collection

of open sets which is a refinement of H. To see that $\sum\sum W_{ij}$ covers D, let p be a point of D and h_β be the first element of H containing p. Then p belongs to some $V_{\beta k}$ but does not belong to $\sum_{\alpha<\beta}\overline{V}_{\alpha k}$. Then for some integer m, p lies in an element of G_m whose closure lies in $V_{\beta k}$ but does not intersect $\sum_{\alpha<\beta}\overline{V}_{\alpha k}$ and p is a point of $U_{\beta k m}$.

For each positive integer k let $X_{k1}, X_{k2}, \ldots$ be a sequence of discrete collections of open sets such that each X_{ki} is a refinement of G_k and $\sum_{i=1}^{\infty} X_{ki}$ covers G^*_k. That S is perfectly screenable follows from the fact that the elements of $\{X_{ki}; i, k = 1, 2, \ldots\}$ may be ordered in a sequence fulfilling the conditions to be satisfied by the sequence $G_1, G_2, \ldots$ mentioned in the definition of a perfectly screenable space.

We find from Example C that condition (b) of Theorem 4 could not be weakened to

(b′) *if p is a point and D is an open set containing p, there is an integer $n(p,D)$ such that an element of $G_n(p, D)$ contains p and lies in D.*

Example C. *A regular strongly screenable topological space not satisfying the first axiom of countability.*[9] Points are the points of the plane. A neighbourhood is either (a) an open interval of a line through the origin such that this interval does not contain the origin or (b) the sum of a collection of open intervals each of which contains the origin and such that each line through the origin contains one of these open intervals.

The space is strongly screenable because for each open covering H of it there are two discrete collections H_1, and H_2 such that $H_1 + H_2$ is an open covering of the space which refines H. The space does not satisfiy the first axiom of countability because for each sequence of neighbourhoods of the origin there is a neighbourhood of the origin that does not contain any of these neighbourhoods. Let G_1 be the collection of all neighbourhoods N of type (b) containing the origin such that if p is a point of the boundary of N, then for some integer n, p is at a distance $1/n$ from the origin in the plane. Then collections $G_2, G_3, \ldots$ may be chosen to satisfy conditions (a) and (b′). However, the space is not metrizable.

2. Developable spaces. For each developable topological space there is a sequence $G_1, G_2, \ldots$ such that (a) G_i is a covering of the space with open sets, (b) G_{i+1} is a refinement of G_i, and (c) for each domain and each point p in D there is an integer $n(p, D)$ such that each element of $G_{n(p,D)}$ which contains p lies in D. Condition (b) is not necessary in defining a developable space because if there is a sequence satisfying conditions (a) and (c), there is one satisfying conditions (a), (b), and (c). In fact, in [5] the condition is imposed that G_{i+1} is a subcollection of G_i. Regular developable topological spaces have been studied extensively because a Moore space is such a space.

[9] A topological space satisfies the first axiom of countability at a point p if there is a countable collection G of neighbourhoods of p such that any neighbourhood of p is a subset of an element of G.

As seen from Examples A and D, not all developable spaces are screenable and not all screenable spaces are developable.

EXAMPLE D. *A regular, separable, strongly screenable space that is not perfectly screenable or developable.* Points belong to the x-axis and neighbourhoods are closed intervals minus their right hand end points.

The space is separable because each set of points dense on the x-axis is dense in the space. If H is an open covering of it, there is an open covering H' which refines H such that no two elements of H' intersect each other. The space is strongly screenable because each such open covering H' is a discrete collection.

The space is not perfectly separable because for each countable collection G of neighbourhoods, there is a point p that does not belong to the left end of any element of G and any neighbourhood of p with a left end at p is not the sum of a subcollection of G. Since the space is separable but not perfectly separable, it is not metrizable. It follows from Theorem 3 that it is not perfectly screenable and from Theorem 5 that it is not developable.

THEOREM 5. *A separable screenable developable space is perfectly separable.*

Proof. No separable space contains uncountably many mutually exclusive domains. Hence if H is an open covering of a separable screenable space, there is a countable open covering H' which refines H. If $G_1, G_2, \ldots$ is a development of a space S and G'_i is a countable open covering that refines G_i, then $\sum G'_i$ is a countable basis for S.

A similar argument shows that a separable perfectly screenable space is perfectly separable.

THEOREM 6. *A strongly screenable developable space is perfectly screenable.*

Proof. Let $G_1, G_2, \ldots$ be a development of the space. Since the space is strongly screenable, for each positive integer i there is a sequence $H_{i1}, H_{i2}, \ldots$ such that H_{ij} is a discrete collection of domains and $\sum_{j=1}^{\infty} H_{ij}$ covers the space and is a refinement of G_i. Then $\{H_{ij}; i, j = 1, 2, \ldots\}$ is a countable collection insuring that the space is perfectly screenable.

THEOREM 7. *A regular developable space (Moore space) is metrizable if it is strongly screenable.*

Theorem 7 follows from Theorems 3 and 6. That Theorem 7 cannot be altered by assuming screenability instead of strong screenability may be seen from Example B.

THEOREM 8. *A screenable Moore space is metrizable if it is normal.*

Proof. This result will follow from Theorem 7 if it is shown that a screenable normal developable space is strongly screenable.

Let H be a collection of mutually exclusive open sets, W be the complement of H^*, and $G_1, G_2, \ldots$ be a development of the space. Denote by X_i the sum of all points p such that no element of G_i containing p intersects W. Since

the space is normal, there is a domain D containing X_i such that $\bar{D}$ does not intersect W. If H_i is the collection of all domains h such that h is the common part of D and an element of H, then H_i is a discrete collection of domains. Since for each collection H of mutually exclusive domains there is a collection $\sum H_i$ covering H such that H_i is a discrete collection of domains which is a refinement of H, a normal developable space is strongly screenable if it is screenable.

THEOREM 9. *For each open covering H of a developable space there is a sequence $X_1, X_2, \ldots$ such that X_i is a discrete collection of closed sets which is a refinement of both X_{i+1} and H while $\sum X_i$ covers the space.*

Proof. Suppose W is a well ordering of H and $G_1, G_2, \ldots$ is a development of the space such that G_{i+1} is a refinement of G_i. For each element h of H, let $x(h, i)$ denote the sum of all points p such that no element of H that contains p precedes h in W and each element of G_i containing p is a subset of h. If X_i denotes the collection of all such sets $x(h, i)$, X_i is a discrete collection because no element of G_i intersects two elements of X_i. If p is a point and $h(p)$ is the first element of H and W containing p, then for some integer i, $[h(p), i]$ contains p. Hence $\sum X_i$ covers the space.

THEOREM 10. *A Moore space is metrizable if it is collectionwise normal.*

Proof. If it is shown that a collectionwise normal Moore space is screenable Theorem 10 will follow from Theorem 8.

For each open covering H of the space, it follows from Theorem 9 that there is a sequence $X_1, X_2, \ldots$ such that X_i is a discrete collection of closed sets and $\sum X_i$ is a covering of the space which is a refinement of H. Collectionwise normality insures that there is a collection Y_i of mutually exclusive open sets covering X^*_i such that no element of Y_i intersects two elements of X_i but each is a subset of an element of H. Then $Y_1, Y_2, \ldots$ is a sequence such that $\sum Y_i$ is a refinement of H covering the space and Y_i is a collection of mutually exclusive domains. Hence, the space is screenable.

Question. Is there a normal Moore space which is neither screenable nor collectionwise normal? If this question could be answered, it could be determined whether or not each normal Moore space is metrizable. F. B. Jones showed [4] that such a space is metrizable if it is separable and $\aleph_1 = C$. Hence, the space mentioned in Example E is metrizable if $\aleph_1 = C$.

EXAMPLE E. *A separable normal Moore space.* The points of the space are the points of the plane which lie above the x-axis and the points of a subset X of the x-axis such that each subset of X is the common part of X and a G_δ set in the plane. The elements of G_i are of two sorts: (a) the interior of a circle of radius less then $1/i$ which lies above the x-axis and (b) $p + I$ where p is a point of X and I is the interior of a circle of diameter less than $1/i$ which is tangent to the x-axis at p from above.

If X is countable, the above space is metrizable because it is perfectly

separable. The set X cannot have the power of the continuum because each subset Y with the power of the continuum in a separable metric space S contains a subset Y' which is not the common part of Y and a G_δ set in S.

In my paper [2] the following theorem is proved.

THEOREM 11. *A topological space is metrizable provided there exists a sequence $H_1, H_2, \ldots$ such that*

(a) *for each integer i, H_i is a collection of sets covering space.*

(b) *a point p is a point of the closure of the set M if and only if for each integer n, some element of H_n contains p and intersects M, and*

(c′) *each pair of points that is covered by either an element of H_{i+1} or the sum of a pair of intersecting elements of H_{i+1} can be covered by an element of H_i.*

In [2] it was falsely stated that (c′) could be replaced by (c) *each pair of points that is covered by the sum of a pair of intersecting elements of H_{i+1} can be covered by an element of H_i.* I am indebted to Dick Wick Hall for calling my attention to the fact that this replacement is not possible. This paper was being studied in one of his classes and L. K. Meals, a member of that class, pointed out that if S is a nonmetrizable space with a development $G_1, G_2, \ldots$ such that G_{i+1} is a refinement of G_i (as in Example A), then if H_i is G_i or S according as i is odd or even, then $H_1, H_2, \ldots$ satisfies conditions (a), (b), and (c).

3. Collectionwise normality. In a developable space, either full normality[10] or collectionwise normality implies metrizability. However, in general, collectionwise normality is weaker than full normality as is shown in the following theorem.

THEOREM 12. *Full normality implies collectionwise normality but not conversely.*

Proof. Let W be a discrete collection of closed sets and H be an open covering of the space such that no element of H intersects two elements of W. If the space is fully normal, there is an open covering H' of the space such that for each point p, the sum of the elements of H' containing p is a subset of an element of H. For each element w of W let D_w be the sum of the elements of H' intersecting w. If w_1 and w_2 are different elements of W, D_{w_1} does not intersect D_{w_2} in a point p or else an element of H containing p intersects both w_1 and w_2. Then the collection of all such domains D_w is a collection of mutually exclusive domains covering W such that no one of these domains intersects two elements of W.

The space described below shows that collectionwise normality does not imply full normality.

[10] A space is fully normal if for each open covering H of the space there is an open covering H' of the space such that the star of each point with respect to H' is a subset of an element of H. Stone showed [6] that the notions of full normality and paracompactness are equivalent for a topological space.

EXAMPLE F. *A collectionwise normal space which is not fully normal.* Points are the elements of an uncountable well ordered collection W such that no element of W is preceded by uncountably many elements of it. A neighbourhood is either the first element of W or the collection of all points that lie between two nonadjacent elements of W.

This space is collectionwise normal because it is normal and does not contain an infinite discrete collection of points. The space is not fully normal because if H is any collection of open sets covering the space, there is a point p such that the star of p with respect to H contains all points that follow p. There is an open covering H such that no elements of H contains all the points that follow some point in it.

THEOREM 13. *Suppose H is an open covering of a collectionwise normal space S and $H_1, H_2, \ldots$ is a sequence such that H_i is a discrete collection of closed sets and $\sum H_i$ is a refinement of H which covers S. Then there is an open covering G of S such that G is a refinement of H and for each point p of an element of H_i there is a domain D containing p such that not more than i elements of G intersect D.*

Since a metric space is developable and collectionwise normal, Theorems 9 and 13 imply that a metric space is paracompact [6]. Theorem 13 would not be true if the hypothesis that $\sum H_i$ covers S were omitted.

Proof of Theorem 13. Since S is collectionwise normal, there is a discrete collection Y_i of open sets covering the sum of the elements of H_i such that Y_i is a refinement of H and each element of Y_i intersects just one element of H_i. As S is normal, there is an open set D_i containing the sum of the elements of H_i such that Y_i covers $\bar{D}_i$. Each element of Y_1 is an element of G and if y is an element of Y_{i+1} not covered by $\bar{D}_1+\bar{D}_2+\ldots+\bar{D}_i$, $y-(\bar{D}_1+\bar{D}_2 \ldots +\bar{D}_i)$ is an element of G.

THEOREM 14. *Collectionwise normality implies normality but normality does not imply collectionwise normality.*

Proof. Since a collection consisting of two mutually exclusive closed sets is a discrete collection, collectionwise normality implies normality. We shall show that the space described in Example G is normal but not collectionwise normal.

EXAMPLE G. *A normal topological space that is not collectionwise normal.* Let P be an uncountable set, Q the set of all subsets of P, and F the set of all functions f on Q having only 1 and 0 as values. To each element p of P associate the function f_p whose value $f_p(q)$ on q is 1 or 0 according as p belongs to q or not. Let F_p be the set of all such functions f_p. The set F is topologized as follows. Any point f in $F - F_p$ is declared to be a neighbourhood of itself. Given a point f_p in F_p and a finite subset r of Q we define the r neighbourhood of f_p to be the set of all f such that $f(q) = f_p(q)$ whenever q belongs to r.

To show the space F thus topologized is normal consider two mutually exclusive closed subsets H_1 and H_2 of it. Let $A_k(k = 1, 2)$ be the set of points common to H_k and F_p and let q_k be the associated set in P consisting of all p for which f_p belongs to A_k. We suppose that neither A_1 nor A_2 is null because if $A_1 = 0$, H_1 and $F - H_1$ are mutually exclusive domains containing H_1 and H_2 respectively. The set D_k of all f in F such that $f(q_k) = 1$ and $f(q_j) = 0$ $(j \neq k)$ is then an open set containing A_k. Moreover no point in F is common both to D_1 and D_2. Therefore the sets $(D_1 - H_2) + (H_1 - A_1)$ and $(D_2 - H_1) + (H_2 - A_2)$ are mutually exclusive open sets containing H_1 and H_2 respectively.

We now show that F is not collectionwise normal. The subset $F_p = \{f_p\}$ of F is a discrete collection of points. However, there does not exist a collection of mutually exclusive neighbourhoods $\{D_p\}$ such that D_p is a neighbourhood of f_p. For suppose, to the contrary, that there were such a collection. Let D_p be the r_p neighbourhood of f_p. Since r_p is a finite subset of Q and P is uncountable there is an integer n and an uncountable subset W of P such that r_p has exactly n elements for every p in W. For any two elements a and b of W the sets r_a and r_b have an element in common, else D_a and D_b would intersect. Hence there is an element q_1 of Q and an uncountable subset W'_1 of W such that q_1 belongs to r_p for every p in W'_1. Moreover there is a t_1 with value 1 or 0 and an uncountable subset W_1 of W'_1 such that $f_p(q_1) = t_1$ for every p in W_1. Similarly there is an element q_2 of Q different from q_1, a t_2 with value 1 or 0, and an uncountable subset W_2 of W_1 such that q_2 belongs to r_2 and $f_p(q_2) = t_2$ for every p in W_2. Continuing recursively in this fashion we get q_k, t_k, W_k for $k = 1, 2, \ldots, n$. Let r be the set consisting of $q_1, q_2, \ldots, q_n$ and D the set of all f with $f(q_k) = t_k$ for $k = 1, 2, \ldots, n$. Then $r_p = r$ and $D_p = D$ for all p in W_n in contradiction to the choice of D_p as mutually exclusive.

One might wonder if Example G could be modified so as to obtain a normal developable space which is not metrizable. A developable space could be obtained by introducing more neighbourhoods into the space F. However, a difficulty might arise in introducing enough neighbourhoods to make the resulting space developable but not enough to make it collectionwise normal.

Another method of modifying Example G is to replace points by closed sets. No non-isolated point of the space F is the intersection of a countable number of neighbourhoods. In Example H, we modify Example G by replacing the points of F by closed sets so as to get a space in which all closed sets are inner limiting (G_δ) sets.[11]

EXAMPLE H. *A normal topological space that is not collectionwise normal and in which each closed set is an inner limiting* (G_δ) *set.* We define P, Q, and F_p as in Example G but let the points of F be functions f defined on Q such

[11] A set is an inner limiting set or G_δ set if it is the intersection of a countable number of open sets.

that $f(x)$ is a non-negative integer. Each point of $F - F_p$ is a neighbourhood. For each finite subset r of Q, each element f_p of F_p, and each positive integer n, f_p plus all points f of F such that $f(x) > n$ (x element of Q) and $f(x') = f_p(x')$ mod 2 (x' element of r) is a neighbourhood.

References

[1] P. Alexandroff and P. Urysohn, *Une condition nécessaire et suffisante pour qu'une classe* (L) *soit une classe* (B), Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences, vol. 177 (1923), 1274-1276.

[2] R. H. Bing, *Extending a Metric*, Duke Math. J., vol. 14 (1947), 511-519.

[3] E. W. Chittenden, *On the metrization problem and related problems in the theory of abstract sets*, Bull. Amer. Math. Soc., vol. 33 (1927), 13-34.

[4] F. B. Jones, *Concerning normal and completely normal spaces*, Bull. Amer. Math. Soc., vol. 43 (1937), 671-679.

[5] R. L. Moore, *Foundations of Point Set Theory*, Amer. Math. Soc. Colloquium Publications, vol. 13, New York, 1932.

[6] A. H. Stone, *Paracompactness and product spaces*, Bull. Amer. Math. Soc., vol. 54 (1948), 977-982.

[7] A. Tychonoff, *Über einen Metrisationssatz von P. Urysohn*, Math. Ann., vol. 95 (1926), 139-142.

[8] P. Urysohn, *Zum Metrisationsproblem*, Math. Ann., vol. 94 (1925), 309-315.

[9] ———, *Über die Mächtigkeit der zusammenhängenden Mengen*, Math. Ann., vol. 94 (1925), 262-295.

[10] G. T. Whyburn, *Analytic Topology*, Amer. Math. Soc. Colloquium Publications, vol. 28, New York, 1942.

University of Virginia and
University of Wisconsin

DISTANCE SETS

L. M. KELLY

1. Introduction. With each set of points S of a distance space there is associated a set of non-negative real munbers $D(S)$ called the distance set of S. The number x is an element of $D(S)$ if and only if x is a distance between some pair of points of S. The number zero is always an element of any distance set and no two distinct elements are equal.

Sierpinski [5], Steinhaus [6], Piccard [4], and many others have considered the relationships existing between S and $D(S)$ for subsets of various spaces, particularly the E_n. Most of these investigations have been concerned with the influence of measure and related properties of S on the associated distance set. For example, it has been shown [6] that the distance set of a set of positive Lebesgue measure must contain an interval with one end point zero. Miss Piccard, on the other hand, has considered the converse problem of investigating the nature of spaces with prescribed distance sets. It is with the sharpening and extending of some of her results and substantial simplification of some of her proofs that this paper is concerned. Theorem 4.2 may be regarded as the principal contribution of this paper, but for the sake of completeness and because of the relative inaccessibility of [4] we have included the simplified proofs of certain basic theorems.

2. Preliminary remarks. If p and q are points of a distance space, the distance between the two elements will be denoted by pq. If P and Q are two subsets, $D(P, Q)$ will represent the set of all distances pq, with $p \in P$ and $q \in Q$. The concept of distance set gives rise to a mapping of the subsets S of a given space onto subsets N of real non-negative numbers, ($D(S) = N$) as well as an inverse mapping of certain subsets of the non-negative numbers on the subsets of the space ($D^{-1}(N) = S$). Of course the inverse mapping need not be, and indeed rarely is, single valued. A subscript will serve to distinguish sets having the same distance set, i.e., $D_s^{-1}(N)$ is a particular set of the space with distance set N.

A sequence of non-negative numbers particularly suited to our purposes is one in which $a_{i+1} > 2a_i$. A finite set of numbers which may be so ordered is called an isosceles set, and an infinite set an isosceles sequence. It is readily seen that any metric space whose distance set is isosceles has all of its triangles isosceles with the base as the shortest side. It is also apparent that any subset of an isosceles sequence or set has the isosceles property.

We proceed now to a consideration of the following questions which seem fundamental in any systematic investigation of distance sets. In what spaces,

Received January 23, 1950.

if any, can an arbitrary set of non-negative numbers including zero be realized as a distance set? What are the corresponding results for denumerable and finite sets? Are there sets of non-negative numbers including zero which cannot be realized in specified spaces, in particular, the Euclidean and Hilbert spaces?

3. Infinite distance sets.

THEOREM 3.1. *An arbitrary set of non-negative numbers including zero is the distance set for some metric space.*

Proof. Let N be such a set and construct a space whose elements are the numbers of N and with distance defined as follows: $pq = \max(p, q)$ if $p \neq q$, and $pq = 0$ if $p = q$. The space is easily seen to be metric with distance set N.

That this result cannot be substantially improved is shown by the following theorem.

THEOREM 3.2. *There exist sets of non-negative real numbers including zero which are not distance sets for any separable metric space.*

Proof. Consider an uncountable set of non-negative numbers, including zero, whose positive numbers are bounded away from zero. Any space with this as distance set is uncountable and discrete, hence not separable.

In order to establish that any countable set of non-negative numbers including zero is a distance set for some subset of Hilbert space, we need the following lemma.

LEMMA 3.1. *If* $0 < a_1 < a_2 \ldots < a_k$, *and*

$$D(k) = \begin{vmatrix} 0 & 1 & 1 & 1 & . & . & 1 \\ 1 & 0 & a_1 & a_2 & . & . & a_k \\ 1 & a_1 & 0 & a_2 & . & . & a_k \\ 1 & a_2 & a_2 & 0 & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & 0 & a_k \\ 1 & a_k & a_k & . & . & a_k & 0 \end{vmatrix}$$

then $\operatorname{sgn} D(k) = (-1)^{k+1}$.

Proof. Let $a_1 \cdot a_2 \cdot a_3 \ldots \cdot a_k = A$. By subtracting appropriate rows and columns, the determinant can be brought into the following form:

$$D(k) = \begin{vmatrix} -2a_1 & a_1 & 0 & . & 0 \\ a_1 & -2a_2 & a_2 & . & 0 \\ 0 & a_2 & -2a_3 & . & 0 \\ . & . & . & . & a_{k-1} \\ 0 & . & . & a_{k-1} & -2a_k \end{vmatrix}$$

$$= -A \cdot \begin{vmatrix} -2 & 1 & 0 & . & 0 \\ a_1/a_2 & -2 & 1 & . & 0 \\ 0 & a_2/a_3 & -2 & . & . \\ . & . & . & -2 & 1 \\ . & . & . & a_{k-1}/a_k & -2 \end{vmatrix} = -A \cdot Q(k).$$

The problem then is one of showing that the sign of the kth order determinant $Q(k)$ is $(-1)^k$. To do this we first establish the fact that $|Q(r)| > |Q(r-1)|$. Noting that the relation is valid for $r = 2$, we assume it true for all $n < r$. The following recursion is easily verified:

$$Q(r) = -2Q(r-1) - \frac{a_{r-1}}{a_r} Q(r-2).$$

Thus $|Q(r)| - |Q(r-1)| = |2Q(r-1) - \frac{a_{r-1}}{a_r} Q(r-2)| - |Q(r-1)|$. If $Q(r-1)$ and $Q(r-2)$ have the same sign, it follows at once that $|Q(r)| > |Q(r-1)|$. If the signs are opposite, we have

$$\begin{aligned}|Q(r)| - |Q(r-1)| &= 2|Q(r-1) - \frac{a_{r-1}}{a_r} Q(r-2)| - |Q(r-1)| \\ &= |Q(r-1)| - \frac{a_{r-1}}{a_r} |Q(r-2)|,\end{aligned}$$

which is greater than zero, since $a_r > a_{r-1}$ and by the inductive hypothesis $|Q(r-1)| > |Q(r-2)|$.

We return now to the problem of establishing the sign of $Q(k)$, assuming that the sign of $Q(r)$ is $(-1)^r$ for $r < k$. Since $Q(k) = -2\,Q(k-1) - \frac{a_{k-1}}{a_k} \cdot Q(k-2)$, it follows that $\operatorname{sgn} Q(k) = -\operatorname{sgn}[2Q(k-1) + \frac{a_{k-1}}{a_k} Q(k-2)]$, and in view of the fact that $|Q(k-1)| > |Q(k-2)|$, we have $\operatorname{sgn} Q(k) = -\operatorname{sgn} Q(k-1) = (-1)^k$. With the observation that the relation is valid for $k = 1, 2$, the induction proof is complete and the lemma is proved.

Remark. It is interesting to note that $Q(k)$ is essentially a continuant. (See any treatise on determinants.)

THEOREM 3.3. *Any countable set of non-negative real numbers including zero is a distance set for some subset of Hilbert space.*

Proof. Let N be such a set and metrize it as in Theorem 3.1. Consider any finite subset of this space consisting of the numbers $0 < a_1 < a_2 < a_3 \ldots < a_k$. The space will be imbeddable congruently in Hilbert space [2, p. 68] if and only if for each k, Δ has the sign $(-1)^{k+1}$ or is zero, where

$$\Delta(k) = \begin{vmatrix} 0 & 1 & 1 & 1 & . & . & 1 \\ 1 & 0 & a_1^2 & a_2^2 & . & . & a_k^2 \\ 1 & a_1^2 & 0 & a_2^2 & . & . & a_k^2 \\ 1 & a_2^2 & a_2^2 & 0 & . & . & . \\ . & . & . & . & . & . & . \\ 1 & . & . & . & . & 0 & a_k^2 \\ 1 & a_k^2 & a_k^2 & . & . & a_k^2 & 0 \end{vmatrix}$$

That Δ has the appropriate sign follows from the lemma.

THEOREM 3.4. *There exist countable sets of non-negative real numbers, including zero, which are not distance sets in any E_n.*

Proof. We will prove that no isosceles sequences $\{a_i\}$ is realizable as a distance set in any E_n. Assume this to be the case for $k < n$ and proceed by induction. Suppose S a subset of E_n with an isosceles sequence as distance set. Let p and q be elements of S such that pq is a minimum. The points of $S - p - q$ are equidistant from p and q and hence are in an E_{n-1}. Furthermore, since S is denumerably infinite, $S - p - q$ is also. If $D(S - p - q)$ were finite, $S - p - q$ would be bounded and hence have an accumulation point. This is impossible and $D(S - p - q)$ is infinite. But $D(S - p - q)$ is a subsequence of an isosceles sequence and is thus itself an isosceles sequence. This contradicts the inductive assumption. To complete the induction, we note that no isosceles sequence is realizable as a distance set in E_1.

Remark. While many other examples might be given of countable sets not realizable as distance sets in any E_n, the following seem particularly worthy of note. The odd integers and zero is not the distance set in any E_n. For Erdös and Anning [1] have shown that any infinite set in E_n all of whose distances are integers is a subset of the line. But it is seen at once that the odd integers and zero cannot be realized as a distance set on the line. As a second example, consider the sequence $1 + 2^{-n}\epsilon$. For "small" ϵ these numbers are "almost equal". But there is no "almost equilateral" infinite set in E_n.

4. Finite distance sets.

THEOREM 4.1. *Any set of n positive numbers and zero is the distance set for a subset of E_n.*

Proof. Let $0 < a_1 < a_2 \ldots < a_n$ be the set of numbers and metrize it as in Theorem 3.1. This space is congruently contained in E_n if and only if the kth ordered bordered principal minors of the determinant $\Delta(n)$ have the sign $(-1)^{k-1}$ or are zero [2, p. 64]. That this is the case follows as in Theorem 3.3.

We are thus assured that any k distinct positive numbers and zero, $k \leqslant n$, can serve as the distance set for some subset of E_n. On the other hand, a set of $\frac{1}{2}n(n - 1) + 1$ "almost equal" positive numbers and zero cannot serve as the distance set for a subset of the E_n since any almost equilateral subset of the E_n consists of at most $n + 1$ points. The question as to whether there exists a set of $n + 1$ positive numbers and zero which is not realizable as a distance set in the E_n naturally arises. In order to establish an affirmative answer to this question we need the following lemmas.

LEMMA 4.1. *If the distance set of a finite metric space is isosceles, the space may be decomposed into two non-null sets P and Q with $D(P, Q) = d$ where d is the diameter of M.*

Proof. Let Q be a maximum set such that $D(p, Q) = d$ where p is an element of $M - Q$, and suppose p' any other element of $M - Q$. From the

triangle inequality it follows that $pq = p'q = d$. Thus Q and $M - Q = P$ is the desired decomposition.

LEMMA 4.2. *If the distance set of a metric space of n points is an isosceles set, it consists of at most n numbers.*

Proof. Proceeding by induction we note that the theorem is true for $n = 1, 2, 3$. Assume it true for all $k < n$. Let P and Q be the two sets of the decomposition of M assured by Lemma 4.1. Suppose P consists of k_1 and Q of k_2 points. Then by the inductive hypothesis, $D(P)$ contains at most k_1 and $D(Q)$ at most k_2 numbers including zero. Thus $D(P)$ and $D(Q)$ together contain at most $k_1 + k_2 - 2 = n - 2$ distinct positive numbers and M has at most $n - 1$ positive numbers.

COROLLARY. *If the distance set of a metric space is an isosceles set of n numbers, the space consists of at least n points.*

DEFINITION. $S_{n,r}$ will denote the surface of the sphere of radius r in E_n.

LEMMA 4.3. *If a set M is a subset of an $S_{n,r}$, but not of any $S_{n-1,r}$, and if the center of $S_{n,r}$ is interior to the convex cover of M, then $r^2 \leqslant \dfrac{n}{2(n+1)} d^2$, where d is the diameter of M.*

Proof. Clearly there are in M vertices of a non-degenerate n-dimensional simplex which contains the centre O of $S_{n,r}$. Using this as a reference simplex, we introduce a barycentric coordinate system and employ a formula of Lagrange. If $m_1, m_2, \ldots, m_{n+1}$ are coordinates of a point Q with $\sum m_i = 1$; $A_1, A_2, \ldots, A_{n+1}$ are vertices of the reference simplex; $a_{ij} = \overline{A_iA_j}$ and P is any other point of E_n, then $\overline{PQ}^2 = \sum \overline{PA}_i^2 m_i - \sum_{i<j} a_{ij}^2 m_i m_j$ (indices from 1 to $n+1$ in all cases). Let $P = Q = O$. Then $0 = r^2 - \sum_{i<j} a_{ij}^2 m_i m_j$, and $r^2 < a^2 \sum_{i<j} m_i m_j$, a the maximum edge. The numbers m_i are all positive since O is interior to the simplex, and it is easily shown that the maximum value of $\sum_{i<j} m_i m_j$ is $n/2(n+1)$, from which it follows that $r^2 \leqslant [n/2(n+1)]a^2 \leqslant [n/2(n+1)]d^2$, and the lemma is proved.

LEMMA 4.4. *If the distance set of a non degenerate n-dimensional simplex is isosceles, the circumcenter is a point of the simplex.*

Proof. Proceeding by induction and noting that the theorem is true for $n = 1, 2$, we assume it true for all $k < n$. Suppose P and Q the sets of the decomposition of the simplex (vertices) assured by Lemma 4.1. The points of P and Q form non degenerate simplices each of dimension less than n and hence the circumcenters O_p and O_q of these "faces" are, by the inductive hypothesis, points of the respective faces. Furthermore, the feet of the perpendiculars from the points of P onto the face determined by Q coincide in O_q

since all the distances pq are equal. Thus O_q is equidistant from the points of P and also from the points of Q, and similarly for O_p. Since O_p and O_q are each in the equidistant locus of the points of P as well as that of the points of Q, the line joining them is also.

Consider now the function px/qx where p is a fixed point of P, q a fixed point of Q, and x a variable point on the closed segment O_pO_q. When $x = O_p$, it follows from Lemma 4.3 and the Pythagorean theorem applied to the triangle pO_pq that $pO_p/qO_p < 1$, while from similar considerations $pO_q/qO_q > 1$ when $x = O_q$. Thus for some point O of the segment O_pO_q, $pO = qO$ and O is the center of the circumsphere of the simplex.

COROLLARY. *If r is the circumradius of a non degenerate n-dimensional simplex with isosceles distance set, then $r^2 < [n/2(n-1)]a^2$ where a is the maximum edge of the simplex.*

THEOREM 4.2. *There exist sets of $n + 1$ positive numbers and zero which are not distance sets for any subset of E_n.*

Proof. Let $\{a_i\} \equiv 0 < a_1 < a_2 \ldots < a_{n+1}$ be an isosceles set of numbers. Proceeding by induction, we will show that such a set is not realizable in E_n. We note that the theorem holds for $n = 1$ and assume then that no set in E_k, $k < n$, can have an isosceles distance set of $k + 2$ numbers.

Suppose S is a set in E_n with $\{a_i\}$ as distance set, and let P and Q be the subsets of the decomposition assured by Lemma 4.1. By the Corollary to Lemma 4.2, S contains at least $n + 2$ points. Suppose the points of P lie irreducibly in an m dimensional subspace. Since the points of Q are equidistant from those of P, they are in an E_{n-m}. By the inductive hypothesis, then $D(P)$ consists of at most m positive numbers and $D(Q)$ at most $n - m$. It follows, since $D(S)$ has n + 1 positive numbers, that $D(P)$ and $D(Q)$ contain exactly m and $n - m$ positive numbers respectively, and that neither contains $D(P, Q) = a$.

The feet of the perpendiculars from the points of Q onto the E_m containing P coincide in a point, say O, equidistant from the points of P. Furthermore, O is equidistant from the points of Q and is in the E_{n-m} containing Q (irreducibly). Thus O is the centre of an m-dimensional sphere containing the points of P and also the centre of an $n - m$ dimensional sphere containing the points of Q. Among the points of P are the vertices of a proper m-dimensional simplex and, by the Corollary to Lemma 4.4, its circumradius is less than the largest edge. Thus $r_p < \frac{1}{2}a$. Similarly, $r_q < \frac{1}{2}a$. But in the triangle pOq, $p \in P$ and $q \in Q$, $pO + Oq \geqslant pq$, that is, $r_p + r_q > a$, a contradiction.

COROLLARY. *There exist sets of $n + k$ numbers, $k = 2, 3, 4, \ldots$, including zero, which are not distance sets for any subset of E_n.*

5. Concluding remarks. Theorem 4.2 focuses attention on sets of points realizable as distance sets in E_{n+1}, but not in E_n. This leads naturally to the definition.

DEFINITION. A set N of positive numbers and zero is said to be ***irreducibly n-dimensional*** relative to Euclidean spaces if it is realizable as a distance set in E_n, but not in E_{n-1}; i.e., $D_x^{-1}(N) \subset E_n$ for some x, but $D_x^{-1}(N) \not\subset E_{n-1}$ for any x.

DEFINITION. A set N is said to be ***properly n-dimensional*** relative to Euclidean spaces if $D_x^{-1}(N) \subset E_n$ for all x and $D_x^{-1}(N) \not\subset E_{n-1}$ for any x.

DEFINITION. A set N is said to be ***rigid*** relative to a space S provided $D_x^{-1}(N) \subset S$ for all x and $D_x^{-1}(N) \subset S$ and $D_y^{-1}(N) \subset S$ implies $D_x^{-1}(N)$ is congruent to $D_y^{-1}(N)$.

Thus any isosceles set of $n+1$ numbers including zero is irreducibly n-dimensional relative to Euclidean spaces, while from the Anning-Erdős theorem, it follows that the even integers, for example, are properly one-dimensional. On the other hand, the distance set of zero together with the integral powers of ten is a D set rigid relative to Euclidean spaces, being realizable on the line in "essentially" only one way. These examples give substance to the definitions, but it would be interesting to know if there exist non-degenerate rigid sets, as well as proper sets, in all dimensions. While we have not yet established the existence of such sets, the following theorem is pertinent.

THEOREM 5.1. *No finite set of non-negative numbers is proper relative to Euclidean spaces.*

Proof. Let N consist of $n+1$ numbers including zero and adjoin to N a second zero, forming the set N^*. Metrize N and N^* as in Theorem 3.1 except for the distance between the two zeros which will be the smallest positive number in N. It is a simple matter to verify that the $(n+2)$-tuple is congruently imbeddable in E_{n+1}, but not E_n. Thus the metrized N and N^* have the same distance sets, but lie in different dimensions.

COROLLARY. *No finite set of non-negative numbers is rigid relative to Euclidean spaces.*

A similar argument is employed to establish the following theorem.

THEOREM 5.2. *No countable set of non-negative numbers whose positive numbers have a minimum is rigid relative to Hilbert space.*

It should be observed that while we have operated largely in Euclidean spaces, many of the results, with obvious modifications, are valid in any locally Euclidean space (i.e., Riemannian), in particular, hyperbolic and elliptic spaces. A more complete analysis of sets of distances peculiar to various familiar spaces is in progress.

References

[1] N. H. Anning and P. Erdös, *Integral distances*, Bull. Amer. Math. Soc., vol. 51 (1945), 598-600.

[2] L. M. Blumenthal, *Distance geometries*, University of Missouri Studies, vol. 13, no. 2 (1938).

[3] P. Erdös, *Integral distances*, Bull. Amer. Math. Soc., vol. 51 (1945), 996.

[4] S. Piccard, *Sur les ensembles de distances des points d'une espace euclidean*. Mem. Univ. Neuchatel, vol. 13, Secretariat de l'Université, Neuchatel, 1939.

[5] W. Sierpinski, *Sur un problème de M. Lusin*. Giornale de Matematiche, 1917.

[6] H. Steinhaus, *Sur les distances des points des ensembles de mesure positive*, Fundamenta Mathematica, vol. 1 (1920), 93-104.

Michigan State College

RADIATION AND GRAVITATIONAL EQUATIONS OF MOTION

L. INFELD AND A. E. SCHEIDEGGER

1. Introduction. Among the classical field theories, general relativity theory occupies a somewhat peculiar place. Unlike those of most other field theories, the field equations in relativity theory are non-linear. This implies that many facts, well known in linear theories, have no analogues in general relativity theory, and conversely. The equations of motion of the sources of the gravitational field are contained in the field equations, a fact which does not apply for the motion of an electron in the electromagnetic field. Conversely, it is difficult to define the notion of a *wave* (familiar in electrodynamics) in relativity theory; for, the linear principle of superposition is crucial for the existence of waves (at least in the sense that the notion of a wave is normally used).

In many other respects, however, close analogues between general relativity theory and other classical field theories exist in spite of the discrepancies mentioned above. It is in part the object of this paper to investigate such analogies.

Since the gravitational field manifests itself in the motion of its sources, the problem of finding the equations of motion is of fundamental importance. This problem was solved some time ago [1], [2]. The general method for obtaining the equations of motion is to introduce an approximation procedure when solving the field equations. Each step of this approximation can be performed only if we impose upon the field certain restrictions (like adding dipoles). These restrictions yield, at the end of the approximation procedure, the differential equations of motion. We have to refer the reader for all the details and also the notation to the paper [2] mentioned above.

Now, in [2] there is at every step of the procedure a certain ambiguity for choosing the solutions of the field equations, which is restricted by assuming a set of co-ordinate conditions. Changing these co-ordinate conditions alters the equations of motion. However, the different equations which can be obtained are physically identical and are merely different mathematical representations of the motion in different co-ordinate systems.

This idea is already partly contained in [2]. There it is shown that rejection of the co-ordinate conditions at one stage of the approximation does not affect the differential equations of motion in the next one [2, §13]. Moreover, at every stage of the approximation the most general solution that can be obtained by rejecting the co-ordinate conditions, is given [2, (9.2) and (13.5)]. However, it is not shown that these solutions can also be obtained by a co-ordinate transformation from the old ones, and are thus equivalent to the old ones.

Received April 19, 1950.

Thus we are led to investigate the influence of co-ordinate transformations upon the equations of motion. This leads, strangely enough, to the destruction of analogies with linear field theories. It is seen that the analogue to the electromagnetic radiation of an accelerated electron exists only formally in general relativity theory. In the case of a gravitational radiation one can perform a co-ordinate transformation and then one regains the solution without radiation (cf. also [3]).

The group of transformations under which the gravitational field equations are covariant, is very general. This implies that the form of the equations of motion depends on the co-ordinates used. There is no physical meaning in the phrase: "the equations of motion of two particles" without reference to the frame in which they apply. It is seen that one can always find a co-ordinate system in which the motion is simply Newtonian. In such a system, however, the metric is very complicated. This is in agreement with a recent statement of Bergmann and Brunings [4] that the co-ordinate system can be chosen so that the equations of motion have *any* form we wish.

Furthermore, it is possible to transfer the whole approximation procedure of [2] into one concerning the co-ordinate system. Thus the condition of integrability of the field equations becomes a condition on the co-ordinate system. We are led to a new version of the usual approximation method: We can enforce integrability not only by adding dipoles but the simple procedure of changing the co-ordinate system.

2. Co-ordinate transformations. We have already mentioned that the general solution of the field equations (by rejecting the co-ordinate conditions) has been calculated [2, (9.2) and (13.5)]. It may be written in the following form:

$$
(2.1)\qquad \begin{cases}
\underset{k-2}{\gamma^*_{00}} = \underset{k-2}{\gamma_{00}} \\
\underset{k-1}{\gamma^*_{0m}} = \underset{k-1}{\gamma_{0m}} + \underset{k-1}{a_{0,m}} \\
\underset{k}{\gamma^*_{mn}} = \underset{k}{\gamma_{mn}} + \underset{k}{a_{m,n}} + \underset{k}{a_{n,m}} - \delta_{mn}\underset{k}{a_{r,r}} + \delta_{mn}\underset{k-1}{a_{0,0}}
\end{cases}
$$

and

$$
(2.2)\qquad \begin{cases}
\underset{k}{\gamma^*_{00}} = \underset{k}{\gamma_{00}} + \underset{k}{b_{r,r}} \\
\underset{k+1}{\gamma^*_{0m}} = \underset{k+1}{\gamma_{0m}} + \underset{k+1}{b_{0,m}} + \underset{k}{b_{m,0}} \\
\underset{k}{\gamma^*_{mn}} = \underset{k}{\gamma_{mn}} + \underset{k}{b_{m,n}} + \underset{k}{b_{n,m}} - \delta_{mn}\underset{k}{b_{r,r}}
\end{cases}
$$

The functions a_0, a_m and b_0, b_m are arbitrary.

We shall investigate whether the general solutions γ^* can be obtained from the particular γ's by a co-ordinate transformation. Let the transformation be

$$x^\beta = x^\beta(x^{\alpha*}) = T^\beta(x^{\alpha*}) \tag{2.3}$$

Then, we calculate the transformed γ's. The transformation of the metric tensor is (assuming the summation convention)

$$g^*_{\alpha\beta} = T^\gamma{}_{|\alpha}\ T^\delta{}_{|\beta}\ g_{\gamma\delta} \tag{2.4}$$

When applying these equations we have to be careful that we take the co-ordinates of the same world point as arguments in all these functions. That is, in addition to the tensorial transformation we have to perform a substitution of the variables x by x^*, according to (2.3).

We develop the tensor $g_{\alpha\beta}$ into a power series with respect to the parameter λ, as this has been done in [2, (5.6)], but now we keep *all* the terms instead of only alternating ones. In the usual solution we assume that the lowest term different from zero is of the order λ^2 (apart from the constant ones $\underset{0}{g}_{\alpha\beta} = \eta_{\alpha\beta}$). We shall confine ourselves to co-ordinate systems where this same property holds. This means that in every co-ordinate system that we admit for consideration, we have a flat Minkowskian metric as a zero approximation of the gravitational field. With this assumption the expanison of the metric tensor becomes

$$\left\{\begin{aligned} g_{mn} &= -\delta_{mn} + \lambda^2 \underset{2}{h}_{mn} + \lambda^3 \underset{3}{h}_{mn} + \dots \\ g_{0m} &= \lambda^2 \underset{2}{h}_{0m} + \lambda^3 \underset{3}{h}_{0m} + \dots \\ g_{00} &= 1 + \lambda^2 \underset{2}{h}_{00} + \lambda^3 \underset{3}{h}_{00} + \dots \end{aligned}\right. \tag{2.5}$$

It was assumed in the original approximation procedure that the motion is "slow", so that one could introduce the "comma-differentiation" [2, (5.3)] with respect to time. If we want to retain in the starred co-ordinate system the assumption that motion is "slow", then we have to assume that also the derivatives of T with respect to x^{0*} are of a higher order in λ than those with respect to x^{m*}; in other words, we have to use the "comma-differentiation" for the transformation function T, too.

Then we can write

$$\left\{\begin{aligned} g^*_{mn} &= T^\alpha{}_{,m}\ T^\beta{}_{,n}\ g_{\alpha\beta} \\ g^*_{m0} &= \lambda\, T^\alpha{}_{,m}\ T^\beta{}_{,0}\ g_{\alpha\beta} \\ g^*_{00} &= \lambda^2 T^\alpha{}_{,0}\ T^\beta{}_{,0}\ g_{\alpha\beta} \end{aligned}\right. \tag{2.6}$$

Equations (2.6) apply quite independently of whether a power development in λ is used for T, or not.

We assume now that the transformation T^α in (2.3) is an infinitesimal co-ordinate transformation, i.e. that it is of the type

$$\left\{\begin{aligned} T^m &= x^{m*} + \lambda^k \underset{(k)}{T^m}(x^{\beta*}) \\ T^0 &= x^{0*} + \lambda^j \underset{(j)}{T^0}(x^{\beta*}) \end{aligned}\right. \tag{2.7}$$

Then, it is easy to see that it is possible to obtain the formulae (2.1) and (2.2) by an appropriate choice of T^β in (2.7). We may indicate the type of calculations involved by taking the following example

$$(2.8)\qquad \begin{cases} x^m = T^m = x^{m*} + \lambda^k \underset{(k)}{T^m}(x^{\beta*}), \\ x^0 = T^0 = x^{0*}. \end{cases}$$

If we insert this together with the expansion of the g's (2.5) into (2.6), we get

$$(2.9)\qquad \begin{cases} g^*_{mn} = g_{mn} + \lambda^k(\delta_{lm}\underset{(k)}{T^s}_{,n} + \delta_{sn}\underset{(k)}{T^l}_{,m})\underset{0}{g_{ls}} + O(k+1) \\ g^*_{m0} = g_{m0} + \underset{(k)}{T^s}_{,0}\,\underset{0}{g_{ms}}\,\lambda^{k+1} \qquad\qquad + O(k+2) \\ g^*_{00} = g_{00} \qquad\qquad\qquad\qquad\qquad + O(k+2) \end{cases}$$

In the above formulae, everything is expressed in the starred co-ordinates. True, one should perform the substitution of the arguments of the occurring functions according to (2.8). This substitution, however, cannot give a contribution to the expressions in (2.9) up to the considered order. Hence we can write in these equations either the starred or the unstarred co-ordinates as arguments. This is particularly so because our zero approximation to the metric tensor in both co-ordinate systems is $\underset{0}{g_{\mu\nu}} = \eta_{\mu\nu}$.

We can express equation (2.9) in terms of the γ's. A straightforward calculation yields

$$(2.10)\qquad \begin{cases} \underset{k}{\gamma^*_{mn}} = \underset{k}{\gamma_{mn}} + \delta_{mn}\underset{k}{T^s}_{,s} - \underset{k}{T^m}_{,n} - \underset{k}{T^n}_{,m} \\ \underset{k}{\gamma^*_{00}} = \underset{k}{\gamma_{00}} - \underset{k}{T^s}_{,s} \\ \underset{k+1}{\gamma^*_{0m}} = \underset{k+1}{\gamma_{0m}} - \underset{k}{T^s}_{,0} \end{cases}$$

This set of equations represents the change of the variables γ under a co-ordinate transformation (2.8). Only the kth and higher approximations are influenced. In a similar way it is seen that the transformation (2.7) with

$$(2.11)\qquad \underset{k}{T^s} = -\underset{k}{a_s}; \qquad \underset{k-1}{T^0} = \underset{k-1}{a_0}$$

yields the expressions (2.1); whereas choosing

$$(2.12)\qquad \underset{k}{T^s} = -\underset{k}{b_s}; \qquad \underset{k+1}{T^0} = \underset{k+1}{b_0}$$

yields (2.2).

These results show that co-ordinate transformations produce all the changes of the γ's which have been found possible in [2] by rejecting the co-ordinate conditions in the kth step of the approximation procedure. Thus, if we have the usual solution, all the different solutions which result from the arbitrariness in the approximation procedure, can be obtained simply by an appropriate co-ordinate transformation, and conversely.

3. Gravitational radiation. It has been shown [5] that the terms omitted in the usual power series for $\gamma_{\alpha\beta}$ in [2, (5.6)] (i.e. the even terms in γ_{0m}, the odd ones in γ_{mn}, γ_{00}) are analogous to the ones representing radiation in electromagnetic theory. We shall adopt here, therefore, the name *radiation terms* for those terms.

One could expect from this analogy that it is possible to deduce similar effects in relativity theory as corresponding to the radiation damping force in electrodynamics. It has been seen [5] that the term which should originate these effects must be of the form

$$\underset{3}{\gamma_{00}} = 0; \quad \underset{4}{\gamma_{0m}} = -4\overset{1}{m}\frac{d^2\eta^m}{d\tau^2} - 4\overset{2}{m}\frac{d^2\zeta^m}{d\tau^2}. \tag{3.1}$$

Starting with this assumption, Hu [6] calculated the equations of motion up to the 9th order. As an illustration he considered two particles of equal mass m moving along circular orbits around each other. The distance r between these two particles may thus be considered as constant up to the order of the Newtonian equations of motion. Then, Hu obtained the result that the total energy defined in Newtonian mechanics as

$$E = \tfrac{1}{2}\,(mv^2 - 2Km^2/r) \tag{3.2}$$

is *increased* by the radiation "damping" force. This result is rather strange from the point of view of Newtonian mechanics, according to which the energy can only be radiated out at the loss of the total energy E.

Our remarks on co-ordinate transformations contained in the last section give the clue for the proper interpretation of Hu's result. It is easily seen that the term (3.1) which was chosen to start the radiation expansion, can be obtained from the usual solution of [2] by putting in (2.8)

$$\underset{3}{T^m} = 4\overset{1}{m}\dot{\eta}^m + 4\overset{2}{m}\dot{\zeta}^m. \tag{3.3}$$

Thus the term starting the radiation expansion is of just such a form that it can be created by a co-ordinate transformation (2.8). Therefore, it also can be wiped out by the corresponding inverse co-ordinate transformation. But in this new co-ordinate system there are no radiation terms, the new metric tensor is that one which we had before the radiation terms were inserted, the equations of motion are the original ones (without the radiation) and thus, we regain the old solution of the relativistic field equations without radiation terms. It may be noted that the co-ordinate system containing the radiation terms with the particular assumption (3.1) does not even require a departure from the usual co-ordinate conditions [2, (9.8)], since $\underset{4}{\gamma_{0m,m}} = \underset{3}{\gamma_{00,0}} = 0$.

We may investigate now whether there are other possibilities for inserting radiation terms. For, generalizing our argument, we are not forced to start radiation terms with the choice of $\underset{4}{\gamma_{0m}}$ as this was done in (3.1). We can ask

whether it is possible to start the omitted terms in the original development for the γ's at any stage of the approximation procedure, say at the $2k$th one.

The prescriptions of [2] imply that one never must add arbitrarily to a field variable any additional poles or higher harmonics, except when this is unavoidable. On the other hand, the equation giving the first radiation terms is one of the following:

$$\underset{2k+1}{\gamma_{00,ss}} = 0; \quad \text{or} \quad \underset{2k}{\gamma_{0m,ss}} = 0; \quad \text{or} \quad \underset{2k+1}{\gamma_{mn,ss}} = 0. \tag{3.4}$$

If we want to take for one of these γ's a solution $\neq 0$ which is nowhere singular in space (including infinity), we see that the only possibility is γ equal to a function of τ. It is readily seen that the particular choice (3.1) suggested by the electromagnetic analogy is indeed of the required form, since η, ζ are functions of τ only. However, a straightforward investigation yields the result that all terms of this type can be created or annihilated by suitable co-ordinate transformations.

One might object to this method that a co-ordinate transformation annihilates the radiation terms only in the lowest approximation where they first appear, but nothing is known as to higher order terms. However, it is possible to carry out the approximation procedure in the new co-ordinate system (without the radiation in the lowest approximation where it was first inserted) and thus to obtain by direct application of the method of [2] the terms originated by the preliminary insertion of the radiation terms and the subsequent co-ordinate transformation. These additional terms could be of two kinds: *either* they are time-functions only and thus may be got rid of by a new co-ordinate transformation of higher order;—*or* they could be singular. If they were singular, this would amount to inserting arbitrarily singular terms at a certain stage in the approximation procedure. The solution of Einstein's field equations would, then, proceed without radiation up to a certain approximation in a suitably chosen co-ordinate system, and then suddenly an additional singular term would be added, which, true enough, could by no means be got rid of and would give a contribution to the equations of motion. However, the approximation procedure, at least as outlined in [2], stands or falls with the prescription that no arbitrary singular terms be inserted at any stage of the procedure. Therefore, if a radiation term should lead to a singular term in higher approximations, after it had been wiped out by a co-ordinate transformation in the approximation where it had first been inserted, it has to be excluded for that very reason. If it leads to additional time-functions only, then those can be annihilated by new (regular!) co-ordinate transformations. Thus, if we add "radiation terms" at a certain stage of the approximation, they are *either* meaningless *or* make the approximation procedure inconsistent.

As already indicated in the introduction to this paper, these results really should have been expected *a priori*. At each step the approximation to the gravitational field variables is only determined up to certain additional terms,

and for the sake of the consistency of the method, it *must* be that the different solutions are changed into each other by mere co-ordinate transformations. As we agreed throughout our work to introduce only such radiation terms which are consistent with all the requirements of the approximation procedure, we *can* really introduce no other solutions than those already found.[1]

4. The equations of motion. We have shown that one can change the relativistic equations of motion in form by performing co-ordinate transformations. In particular, we were able in the last section to create and annihilate radiation terms in the equations of motion as well as in the expansions for the field variables.

Now, we may ask: What is generally the influence of a co-ordinate transformation of the type (2.7) upon the relativistic equations of motion? Is it perhaps possible to adjust the co-ordinate system at every step of the approximation procedure in such a way that the motion has always a certain standard form? Intuitively, one could expect that a co-ordinate transformation can change the equations of motion to any form we like. However, considering only infinitesimal transformations, we cannot *a priori* be sure that this is true.

We have already seen before that radiation terms are irrelevant. Thus, we may as well stick to the power series [2, (5.6)]. In order to apply a co-ordinate transformation we assume that the field equations are solved up to the order $2k + 1$. Then we know the following quantities:

$$\underset{2}{\gamma_{00}} \cdots \underset{2k}{\gamma_{00}};\ \underset{3}{\gamma_{0m}} \cdots \underset{2k+1}{\gamma_{0m}};\ \underset{4}{\gamma_{mn}} \cdots \underset{2k}{\gamma_{mn}} \tag{4.1}$$

and the equations of motion of the corresponding order are

$$\lambda^4 \underset{4}{\overset{i}{C}}_m(\eta, \zeta) + \ldots + \lambda^{2k} \underset{(2k)}{\overset{i}{C}}_m(\eta, \zeta) = 0 \qquad (i = 1, 2). \tag{4.2}$$

After this step we consider two cases. In the first one we go on in the usual manner of [2], but in the second case we perform an infinitesimal co-ordinate transformation.

Thus in the *first* case we shall calculate the field variables in the old co-ordinate system of [2] up to the $(2k + 4)$th order, and similarly we proceed with the equations of motion. The latter will be

$$\lambda^4 \underset{4}{\overset{i}{C}}_m(\eta, \zeta) + \ldots + \lambda^{2k+4} \underset{2k+4}{\overset{i}{C}}_m(\eta, \zeta) = 0 \qquad (i = 1, 2). \tag{4.3}$$

In the *second* case, we proceed in a different way. We perform an infinitesimal transformation before going on with the approximation procedure:

$$x^m = x^{m*} + \lambda^{2k} \underset{(2k)}{T^m}(x^*). \tag{4.4}$$

[1] A more elaborate discussion of this subject by means of a slightly different approach has been given earlier [3].

This changes the values of the field variables (4.1) according to (2.12), where everything is now expressed in x^*. As was shown before, the effect of this co-ordinate transformation (4.4) is the same as choosing different solutions of the field equations at the step *before* obtaining (4.1). Thus, in the "new" co-ordinate system we have, up to the order $2k$, the following expressions for the field variables and equations of motion:

$$(4.5)\qquad \begin{cases} \underset{2}{\gamma_{00}} \ldots \underset{2k-2}{\gamma_{00}},\ \underset{2k}{\gamma_{00}} - \underset{2k}{T^r}_{,r} \\ \underset{3}{\gamma_{0m}} \ldots \underset{2k-1}{\gamma_{0m}},\ \underset{2k+1}{\gamma_{0m}} - \underset{2k}{T^m}_{,0} \\ \underset{4}{\gamma_{mn}} \ldots \underset{2k-2}{\gamma_{mn}},\ \underset{2k}{\gamma_{mn}} - \underset{2k}{T^m}_{,n} - \underset{2k}{T^n}_{,m} + \delta_{mn}\underset{2k}{T^s}_{,s} \end{cases}$$

$$(4.6)\qquad \lambda^4 \overset{i}{\underset{4}{C}}(\eta^*\zeta^*) + \ldots + \lambda^{2k}\overset{i}{\underset{(2k)}{C}}(\eta^*\zeta^*) = 0 \qquad (i = 1, 2).$$

In the above equations it is understood that one has to replace the original arguments x in all the occurring functions by x^* (and hence also η, ζ by η^*, ζ^* respectively).

With the values (4.5) for the field variables we can go on with the approximation procedure as in the first case. After performing two more steps in the approximation method we obtain the equations of motion of the order $2k + 4$, now expressed entirely in the *new* co-ordinates. It is to be expected that they will be formally different from those obtained by the procedure performed in the first case above.

To investigate this question we calculate the new equations of motion after taking the new solutions (4.5) for the γ's up to the $2k$th step. Since one has to go through two stages of the approximation method, this is quite a laborious undertaking.

We can simplify the computational work involved by making some special assumptions. We may note that we need the behaviour of the transformation (4.4) in any case only in the neighbourhood of the world-lines of the particles. Thus we can develop the expression for $\underset{2k}{T^\mu}$ around the world-lines into a Taylor series. Herein we assume that the first and second space derivatives shall vanish. Moreover, we assume that only $\underset{2k}{T^m}$ is different from zero, whereas $\underset{2k}{T^0}$ vanishes. Thus we have, near the first world-line, the following co-ordinate transformation[2]

$$(4.7)\qquad x^m = x^{m*} + \lambda^{2k}\overset{1}{\underset{(2k)}{T^m}}(x^*).$$

Then, the only γ which is influenced up to the order $2k + 1$ is γ_{0m}. It becomes, according to (2.10),

$$(4.8)\qquad \underset{2k+1}{\gamma^*_{0m}} = \underset{2k+1}{\gamma_{0m}} - \overset{1}{\underset{2k}{T^m}}_{,0}.$$

[2] The index "1" above T means that this transformation is different from the identity only in the neighbourhood of the *first* world-line.

When proceeding to the higher approximations, we are interested only in terms that contain $\underset{2k}{T}{}^m$; the remaining ones are just those which we would have got without the co-ordinate transformation. Keeping only terms that contain T, we obtain the difference between the equations of motion in the new and in the old co-ordinate systems. We may note that we can use the standard co-ordinate conditions [2, (9.8)] throughout; for, (4.8) satisfies these conditions in the neighbourhood of the first world-line and for the subsequent steps we are free to choose any co-ordinate conditions we like.

Starting our calculations, we have first to compute $\underset{2k+3}{\Lambda^*_{0n}}$. The result is

$$\underset{2k+3}{\Lambda^*_{0m}} = \underset{2k+3}{\Lambda_{0m}} + \tfrac{1}{2}\,\phi_{,ms}\,\underset{2k}{\overset{1}{T}}{}^{s}{}_{,0}. \tag{4.9}$$

We may note that the expression (4.9) is obtained from calculations made in [2]. For, we observe that in our problem $\underset{2k}{\overset{1}{T}}{}^m$ can only combine with terms of the order *two* so as to yield expressions of the order $2k + 3$. Hence it is seen that we obtain in $\underset{2k+3}{\Lambda^*_{0m}}$ the same contribution from $\underset{2k}{\overset{1}{T}}{}^m$ as we have in $\underset{5}{\Lambda_{0m}}$ from $-\underset{3}{\gamma_{0m}}$. Thus the result (4.9) is obtained immediately from [2, (A.5.2)].

To find now $\underset{2k+3}{\gamma_{0m}}$ we can, of course, use the standard co-ordinate conditions. Then, we obtain near the first world-line

$$\underset{2k+3}{\gamma^*_{0n}} = \underset{2k+3}{\gamma_{0n}} + \tfrac{1}{2}\,\phi_{,n}\,\underset{2k}{\overset{1}{T}}{}^{l}{}_{,0}\,\{x^l - \eta^l\} - \tfrac{1}{2}\,\phi\,\underset{2k}{\overset{1}{T}}{}^{n}{}_{,0}. \tag{4.10}$$

The next step is to calculate $\underset{2k+4}{\Lambda^*_{mn}}$. We obtain it in a similar way as $\underset{2k+3}{\Lambda^*_{0m}}$ above from the calculations in [2]. Using [2, (A.12.3)], we get the following result:

$$\left\{\begin{aligned} 2\underset{2k+4}{\Lambda^*_{mn}} = {} & 2\underset{2k+4}{\Lambda_{mn}} - \phi_{,0mn}\,\underset{2k}{\overset{1}{T}}{}^{l}{}_{,0}\,\{x^l - \eta^l\} - \underline{\phi_{,mn}\,\underset{2k}{\overset{1}{T}}{}^{l}{}_{,00}\,\{x^l - \eta^l\}} \\ & - 2\underset{2k}{\overset{1}{T}}{}^{l}{}_{,0}\;\underset{3}{\gamma_{0l,mn}} + \underset{2k}{\overset{1}{T}}{}^{l}{}_{,0}\;\underset{3}{\gamma_{0n,ml}} + \underset{2k}{\overset{1}{T}}{}^{m}{}_{,0}\;\phi_{,0n} \\ & + \phi_{,0m}\underset{2k}{\overset{1}{T}}{}^{n}{}_{,0} - 2\delta_{mn}\underset{2k}{\overset{1}{T}}{}^{l}{}_{,0}\;\phi_{,0l} + \underline{\phi_{,m}\,\underset{2k}{\overset{1}{T}}{}_{,00}} \\ & + \underline{\phi_{,n}\,\underset{2k}{\overset{1}{T}}{}^{m}{}_{,00}} + \underset{2k}{\overset{1}{T}}{}^{s}{}_{,0}\;\underset{3}{\gamma_{0m,ns}} + \phi_{,mn}\,\underset{2k}{\overset{1}{T}}{}^{l}{}_{,0}\;\dot{\eta}^l. \end{aligned}\right. \tag{4.11}$$

The next step is to find the surface integrals

$$\underset{2k+4}{\overset{1}{C}}{}_m = \int \underset{2k+4}{\Lambda^*_{mn}}\,n_n\,ds. \tag{4.12}$$

Only the underlined terms in (4.11) can give a contribution, since only these approach infinity like $1/r^2$ near the first particle. Evaluating these integrals yields the following result:

$$\underset{2k+4}{\overset{1}{C^*_m}} = \underset{2k+4}{\overset{1}{C_m}} + 4\underset{2k}{\overset{1}{T^m}}{}_{,00}\,\overset{1}{m}. \tag{4.13}$$

Thus, the equations of motion have in the new co-ordinate system the following form:

$$\lambda^4 \underset{4}{\overset{1}{C_m}} + \ldots + \lambda^{2k+4}\{\underset{2k+4}{\overset{1}{C_m}} + 4\underset{2k}{\overset{1}{T^m}}{}_{,00}\,\overset{1}{m}\} = 0. \tag{4.14}$$

Our work above was referring to one world-line only. However, we can arrange very well that our transformation term $\underset{2k}{\overset{1}{T^m}}$ is zero near the second world-line so as not to influence the surface integrals around the second particle. Conversely, we can assume another transformation $\underset{2k}{\overset{2}{T^m}}$ of a type similar to the first one, which changes the surface integrals for the second particle only. Then, both transformations together yield the following equations of motion of the order $2k + 4$:

$$\lambda^4 \underset{4}{\overset{i}{C_m}} + \ldots + \lambda^{2k+4}[\underset{2k+4}{\overset{i}{C_m}} + 4\underset{2k}{\overset{i}{T^m}}{}_{,00}\,\overset{i}{m}] = 0 \qquad (i = 1, 2). \tag{4.15}$$

The old functions η, ζ are functions of time only. Hence it is seen that it is always possible to choose transformations so that the square brackets vanish. This means that we can always choose a transformation $\underset{2k}{T^s}$ so that the equations of motion of the order $2k + 4$ do not contain any terms of that order at all. Thus we see that our restrictive assumptions for the admissible co-ordinate transformations are still sufficiently wide to allow us to construct a co-ordinate system where the co-efficient of λ^{2k} in the equations of motion of the order $2j > 2k$ vanishes.

This argument can be repeated. Let us assume that we have solved the field equations up to the order $2j > 2$ and obtained the equations of motion of the same order. Then, we can preform a co-ordinate transformation choosing $\underset{2}{\overset{1,2}{T^m}}$ so that the terms connected with the power λ^6 vanish in the new equations of motion. That will change all of the $\underset{2k}{C_m}$'s with $k<j$. Then, we transform the equations of motion again, choosing $\underset{4}{T}$ so that the new $\underset{8}{C}$'s in that new co-ordinate system vanish, etc. Finally, we shall end up with differential equations of motion of the $2j$th order, but containing only $\underset{4}{C_m}$. This means that it is always possible to construct a co-ordinate system so that the equations of motion are just Newtonian.[3]

[3] A detailed investigation of all these statements may be found in [8].

This is the standard form to which the differential equations of motion can be reduced. We cannot go further and reduce e.g. the two particles to being at rest with respect to each other, because we have explicitly assumed that our admissible co-ordinate transformations be different from the identity transformation at most by a term proportional to λ^2.

It is also possible to arrive at the same result by expressing η, ζ in (4.3) in the new system by means of (4.4) instead of calculating the equations of motion anew in the transformed co-ordinate system.

5. Conclusions. We have seen that it is always possible to set up such a co-ordinate system that the relativistic equations of motion of any order have Newtonian form. We shall now investigate what conclusions can be drawn from this statement.

First of all, we have to emphasize that our foregoing mathematical deductions do *not* imply that the motion is the same as it would be unrelativistically in such a specially chosen co-ordinate system. Only the *form* of the differential equations of motion is Newtonian; we must keep in mind, however, that the metric in this case is by no means of Newtonian character near the singularities. If we wish that the metric field be of Newtonian character near its sources, then the motion is non-Newtonian and the equations of motion are as calculated in [2].

So far, this does not yield any new ideas. We may note, however, that the above statement about the possible Newtonian form of the equations of motion can be formulated in a slightly different way. For, we observe that it is the same thing as saying that, at every step of the approximation procedure, we can reach the vanishing of the corresponding surface integrals by choosing the co-ordinate system two steps before in an appropriate way. This shows that we really have found a new version of the method in [2] for solving Einstein's field equations, which is equivalent to the one introducing and annihilating dipoles.

Let us formulate this conception somewhat more precisely. Assume that the field equations are to be solved by making the usual expansion [2, (5.6)] of the field variables with respect to the parameter λ. Suppose the field equations have been solved up to a certain stage. Proceeding one step further, we are faced with the task of solving the following system of equations:

$$\underset{2k-2}{\Phi_{00}} + 2 \underset{2k-2}{\Lambda_{00}} = 0 \tag{5.1a}$$

$$\underset{2k-1}{\Phi_{0m}} + 2 \underset{2k-1}{\Lambda_{0m}} = 0 \tag{5.1b}$$

$$\underset{2k}{\Phi_{mn}} + 2 \underset{2k}{\Lambda_{mn}} = 0 \tag{5.1c}$$

(cf. [2, (8.1)]). Because of the Bianchi identities this system is generally not solvable. In order to solve (5.1c) we have to add dipoles to the *known* solution $\underset{2k-2}{\gamma_{00}}$. For, then we can obtain that the surface integrals

$$\underset{2k-1}{\overset{i}{C}}{}_0 = \frac{1}{4\pi}\int_i \underset{2k-1}{\Lambda_{0n}}\, n_n\, dS \tag{5.2}$$

$$\underset{2k}{\overset{i}{C}}{}_m = \frac{1}{4\pi}\int_i \underset{2k}{\Lambda_{mn}}\, n_n\, dS \tag{5.3}$$

vanish, which is the condition of integrability. Indeed, by inspecting (5.3) we note that adding dipoles $\underset{2k-2}{\overset{i}{S}}{}_m\psi_{,m}$ to $\underset{2k-2}{\gamma_{00}}$ changes these surface integrals into $\underset{2k}{\overset{i}{C^*}}{}_m$ with

$$\underset{2k}{\overset{i}{C^*}}{}_m = \underset{2k}{\overset{i}{C}}{}_m + \underset{2k-2}{\overset{i}{S}}{}_{m,\,00} \tag{5.4}$$

which can be made zero by choosing

$$\underset{2k-2}{\overset{i}{S}}{}_{m,\,00} = -\underset{2k}{\overset{i}{C}}{}_m. \tag{5.5}$$

However, the contribution to the surface integrals obtained by adding dipoles to $\underset{2k-2}{\gamma_{00}}$ is very similar to the one obtained by performing a co-ordinate transformation at the $(2k - 4)$th stage of the approximation procedure. We have seen that a co-ordinate transformation

$$x^m = x^{m*} + \lambda^{2k-4} \underset{(2k-4)}{T^m}(x^*) \tag{5.6}$$

causes a change in the surface integrals. This change is, if we assume that the space derivatives of T vanish,

$$\underset{2k}{\overset{i}{C^*}}{}_m = \underset{2k}{\overset{i}{C}}{}_m(\eta^*, \zeta^*) + 4\overset{i}{m}\, \underset{2k-4}{\overset{i}{T}}{}^{m}{}_{,00}(x^* = \eta^*) \tag{5.7}$$

Thus, to enforce the integrability of (5.1c) we can either add dipoles or change the co-ordinate system according to (5.6). Furthermore, we see that $\underset{2k-4}{S}{}_{m,\,00}$ in $\underset{2k-2}{\gamma_{00}}$ has the same effect upon the equations of motion as $4m\underset{2k-4}{T^m}$ in the $(2k - 4)$th step of the approximation.

We may emphasize once more that the form in which the equations of motion finally appear does not influence any of the well known results of general relativity theory. It is only a matter of representation whether these relativistic effects are explicitly contained in the equations of motion or in the metric field.

Let us illustrate this by a specific example. Robertson [7] has integrated the differential relativistic equations of motion of the sixth order for the two-body problem. His result was that one obtains the same effect as when considering the motion of a small body in the Schwarzschild field of a large one by applying the geodesic principle; i.e., the orbit of a double star in general relativity theory

differs in its secular behaviour from the classical orbit only in an advance of perihelion equal to that which an infinitesimal planet, describing the same relative orbit, would undergo in the field of a star whose mass is the sum of those of the two components of the double star. Hence it is intuitively seen that introducing a co-ordinate system rotating at the right speed will reduce the non-Newtonian orbit to a Newtonian one. This co-ordinate transformation needs only to take place in the immediate neighbourhood of the trajectory, a statement which is in agreement with the fact that we had to know T^m only near the world lines of the particles. A detailed investigation (in [8]) shows indeed that it is possible to find such a co-ordinate system and also that the metric in that system contains components $g_{0m} \neq 0$ which confirms that light rays no longer have a simple trajectory in that system.

Thus we obtain either simple (Newtonian) equations of motion and a complicated metric field, or a simple field (of Newtonian character near the singularities) but non-Newtonian equations of motion.

References

[1] A. Einstein, L. Infeld and B. Hoffmann, *The gravitational equations and the problem of motion*, Ann. of Math., vol. 39 (1938), 66-100.

[2] A. Einstein and L. Infeld, *On the motion of particles in general relativity theory*, Can. J. Math., vol. 1 (1949), 209-241.

[3] A. E. Scheidegger, *On gravitational radiation*, Proceedings of the Second Canadian Mathematical Congress, Vancouver 1949, in press.

[4] P. G. Bergmann and Johanna Brunings, *Non linear field theories II; canonical equations and quantisation*, Reviews of Modern Physics, vol. 21 (1949), 480-487.

[5] L. Infeld, *Electromagnetic and gravitational radiation*, Physical Review, vol. 53 (1938), 836-841.

[6] N. Hu, *Radiation damping in the general theory of relativity*, Proc. Roy. Irish Acad., vol. 51A (1947), 87-111.

[7] H. P. Robertson, *The two body problem in general relativity*, Ann. of Math., vol. 39 (1938), 101-104.

[8] A. E. Scheidegger, *Radiation and gravitational equations of motion*, Thesis, University of Toronto Library, 1950.

Department of Applied Mathematics,
University of Toronto

NON-NORMAL GALOIS THEORY FOR NON-COMMUTATIVE AND NON-SEMISIMPLE RINGS

TADASI NAKAYAMA

THE purpose of the present work is to give, as a continuation of the writer's study of Galois theory for general rings ([8], [9], [10]), a kind of Galois theory for general, non-commutative and non-semisimple rings, which includes, at least in its main features, the Kaloujnine-Jacobson Galois theory of non-normal fields ([3]; cf. [4], [5]). To deal with the non-commutativity we bring to the fore certain double-moduli rather than self-composites, while the non-semisimplicity is manipulated by the method and idea used in the writer's above mentioned study on (normal) Galois theory and commuter systems of non-semisimple rings. (For the normal Galois theory of rings cf. [1], [2], [6], [7], [11], besides the above.) Some of our arguments may even serve to make some simplification in Jacobson's treatment of ordinary fields.

1. Galois ring and Galois system of module-endomorphisms of a ring. Throughout this paper a ring means a ring with unit element and its module, right or left, one for which the unit element is the identity operator.

Let A be a semiprimary ring. An A-right-, say, module $\mathfrak{m}$ is called regular when a direct sum of a certain (finite) number, say v, of its copies is (A-) isomorphic to the direct sum of a certain number, say u, of copies of the A-right-module A. The A-endomorphism ring A^* of $\mathfrak{m}$ is nothing but the commuter ring of A in the absolute endomorphism ring of $\mathfrak{m}$. We have

LEMMA. *The number u/v,* called the (A-) rank of the regular module $\mathfrak{m}$, *is determined uniquely and characterizes the structure of* $\mathfrak{m}$. *A^* is semiprimary and* $\mathfrak{m}$ *is also regular with respect to A^*. The A^*-rank of* $\mathfrak{m}$ *is inverse to the A-rank. The A^*-endomorphism ring of* $\mathfrak{m}$ *coincides with A.*

If $\mathfrak{m}$ is also regular with respect to a (semiprimary) subring B, then any other regular A-right-module $\mathfrak{n}$ is regular with respect to B too and the ratio of the A-ranks of $\mathfrak{m}$, $\mathfrak{n}$ is equal to that of their B-ranks.

We note further that if A satisfies the minimum condition for right-ideals then A^* satisfies the same for left-ideals.

Throughout this paper, R will denote a ring satisfying the minimum con-

Received December 1, 1949, revised March 17, 1950. *Addendum in revision.* After the submission of the present paper to the Journal, the writer had access to a paper by G. Hochschild, entitled "Double vector spaces over division rings" (Amer. Jour. Math., vol. 71 (1949)), closely related to the present one. The idea of considering in the non-commutative case certain double-moduli, rather than self-composites, has been exploited there already. However, in the present work we dealt with non-division rings, in fact with general non-semisimple rings.

dition for left-ideals.[1] Let **A** be its absolute endomorphism ring, that is, the endomorphism ring of R as module without operator domain. Denote by R_l, R_r, or generally X_l, X_r with a subset X of R, the set of left-, right-multiplications of R, or X, upon R. Let **B** be a subring of **A**, i.e. a certain ring of (module-)endomorphisms of R, which contains R_l. When further the direct sum of a certain number, say s, of copies of R is (**B**-) isomorphic to the **B**-right-module **B** itself, i.e., when R is **B**-regular with s^{-1} as rank, we call **B** a *Galois ring* of module-endomorphisms of R. Here **B** satisfies the minimum condition for R_l-right-submoduli, whence certainly that for its right-ideals. The commuter ring $V(\mathbf{B})$ of **B** in **A**, i.e. the **B**-endomorphism ring of R, is contained in $V(R_l) = R_r$, and so has a form S_r with a subring S of R. R is S_r-regular with rank s, that is, R has an independent S-right-basis of s terms. Moreover, $V(S_r) = \mathbf{B}$.

If conversely S is a subring of R such that R possesses an independent (finite) S-right-basis, of s terms, say, then S certainly satisfies, together with R, the minimum condition for left-ideals[2] and $V(S_r) = \mathbf{B}$ is a Galois ring in the above sense. And $S_r = V(\mathbf{B})$. Thus

THEOREM 1. *Galois rings* **B** *of module-endomorphisms and subrings S such that R has independent right-basis over S are in* 1-1 *dual correspondence, by* $V(\mathbf{B}) = S_r$, $\mathbf{B} = V(S_r)$. *The* **B**-*rank of R is inverse to the S_r- (that is, S-right-) of R.*

Further, by the Lemma, applied to the **B**-, and R_l-module **B**, instead of A-, and B-module $\mathfrak{n}$, we see that **B** is R_l (right-) regular and the R_l-rank of **B** is equal to the S-rank s of R. Hence

THEOREM 2. *The Galois ring* **B** *has an independent right-basis of s terms over its subring R_l, where* $1/s$ *is the* **B**-*rank of R (that is, s is the S_r-rank of R):*

$$\mathbf{B} = \beta_1 R_l \oplus \beta_2 R_l \oplus \ldots \oplus \beta s R_l.$$

We call such an independent right-basis of a Galois ring over R a *Galois system* of module-endomorphisms of R. Our next task will then be the construction of such a Galois system.

2. Construction of Galois system. Let $\mathfrak{m}$ be a right-module of R and $\mathfrak{n}$ be a left-module of R. By their direct product $\mathfrak{m} \times \mathfrak{n} = \mathfrak{m} \times_R \mathfrak{n}$ we mean, as usual, a module generated (freely) by symbols uv ($u \in \mathfrak{m}$, $v \in \mathfrak{n}$) with relations

$$(u_1 + u_2)v = u_1v + u_2v, \qquad u(v_1 + v_2) = uv_1 + uv_2,$$
$$(uz)v = u(zv) \quad (z \in R).$$

[1]We can develop our whole theory also under the assumpton that R, B, S (see below) are semiprimary rings, or even under a much weaker assumption as G. Azumaya has kindly pointed out. However, the writer prefers to present the theory in the form below where R (and then S) satisfies the minimum condition, since the assumption does not spoil the essential feature of the theory.

[2]Consider $R\mathfrak{l}$ with left-ideals $\mathfrak{l}$ of S.

If $\mathfrak{n}$ is an R-double-module the product $\mathfrak{m} \times \mathfrak{n}$ is, in a natural manner, an R-right-module, and if both $\mathfrak{m}$, $\mathfrak{n}$ are R-double-modules then $\mathfrak{m} \times \mathfrak{n}$ becomes an R-double-module. In case $\mathfrak{m}$ possesses an independent R-right-basis $(u_1, u_2, \dots, u_m)$ we have $\mathfrak{m} \times \mathfrak{n} = u_1\mathfrak{n} \oplus u_2\mathfrak{n} \oplus \dots \oplus u_m\mathfrak{n}$. If also $\mathfrak{n}$ possesses an independent R-left-basis $(v_1, v_2, \dots, v_n)$, then $\mathfrak{m} \times \mathfrak{n} = \sum^0_{i,j} u_i R v_j$.

Now, let S be a subring of R such that R has an independent S-right-basis of s, say, terms. We consider R as S-right-, and S-left-module, and we want to construct the direct self-product $R \times R$ over S. However, to avoid ambiguity in notation, we introduce two (ring-) isomorphisms σ, τ of R. Putting $zx^\sigma = (zx)^\sigma$, $x^\sigma z = (xz)^\sigma$, $x^\tau z = (xz)^\tau$, $zx^\tau = (zx)^\tau$ $(x, z \in R)$ we consider R^σ, R^τ as R-double-moduli. We then construct

$$(1) \qquad R^\sigma \times R^\tau = R^\sigma \times_S R^\tau = x_1^\sigma R^\tau \oplus x_2^\sigma R^\tau \oplus \dots \oplus x_s^\sigma R^\tau,$$

where $(x_1, x_2, \dots, x_s)$ is an independent S-right-basis of R.

According to (1) we have, for each $z \in R$,

$$(2) \qquad z^\sigma 1^\tau = x_1^\sigma \beta_1(z)^\tau + x_2^\sigma \beta_2(z)^\tau + \dots + x_s^\sigma \beta_s(z)^\tau, \qquad \beta_h(z) \in R.$$

$\beta_h(z)$ are determined uniquely by z, and β_h: $z \to \beta_h(z)$ $(h = 1, 2, \dots, s)$ are module-endomorphisms of R. Moreover, for $a \in S$ we have $(za)^\sigma 1^\tau = z^\sigma a^\tau = \sum_h x_h^\sigma (\beta_h(z)a)^\tau$. Thus $\beta_h(z)a = \beta_h(za)$, or $\beta_h a_r = a_r \beta_h$, and so β_h are S_r-endomorphisms of R, and $\beta_h \in V(S_r) = \mathbf{B}$. We assert that they form a Galois system belonging to S. Observe first that

$$x_i^\sigma 1^\tau = x_1^\sigma \beta_1(x_i)^\tau + x_2^\sigma \beta_2(x_i) + \dots + x_s^\sigma \beta_s(x_i)$$

and so
$$\beta_h(x_i) = \delta_{hi} \qquad \text{(Kronecker } \delta\text{)}.$$

Therefore $\beta_h y_l$, with $y \in R$, maps x_i upon $y\delta_{hi}$, and $\sum_h \beta_h y_{hl}$ $(y_h \in R)$ maps x_i upon y_i. It follows that $\beta_1, \beta_2, \dots, \beta_s$ are R_l-right-independent. Moreover, since y_h may be taken arbitrarily, the totality of $\sum_h \beta_h y_{hl}$ coincides with the whole $V(S_r) = \mathbf{B}$;

$$(3) \qquad V(S_r) = \mathbf{B} = \beta_1 R_l \oplus \beta_2 R_l \oplus \dots \oplus \beta_s R_l.$$

If $z = \sum_h x_h a_h$ $(a_h \in S)$ then $\beta_h(z) = a_h$. Thus

THEOREM 3. *The s (module-)endomorphisms β_h of R defined in* (2) *form a Galois system belonging to the subring S. Here $\beta_h(R) = S$ for each h. Moreover $\beta_h(z) = 0$ $(h = 1, 2, \dots, s)$ (that is, $z^\sigma 1^\tau = 0$) implies $z = 0$.*

3. Double-moduli and their relation moduli. Although Theorems 1, 2, 3 already give the main features of our Galois theory, it is useful as well as important to extend the above construction of a Galois system to the case of general double-moduli of a certain type and thus obtain a characterization of a Galois ring (Theorem 7). It is our purpose to generalize Jacobson's theory of self-composites of (commutative) fields, but we have to adopt a somewhat

different formulation and method, because of the non-commutativity and the non-semisimplicity of R.

Let $\mathfrak{M}$ be a double-module of R having an independent R-right-basis, and let u_0 be an element of $\mathfrak{M}$. Let $(u_1, u_2, \ldots, u_m)$ be an independent R-right-basis of $\mathfrak{M}$, and put

(4) $$zu_0 = u_1\mu_1(z) + u_2\mu_2(z) + \ldots + u_m\mu_m(z)$$

for $z \in R$; $\mu_1, \mu_2, \ldots, \mu_m$ are module-endomorphisms of R.

Consider further a second R-double-module $\mathfrak{N}$ with an independent R-right-basis $(v_1, v_2, \ldots, v_n)$, and its element v_0. Introduce module-endomorphisms $\nu_1, \nu_2, \ldots, \nu_n$ of R correspondingly by

(5) $$zv_0 = v_1\nu_1(z) + v_2\nu_2(z) + \ldots + v_n\nu_n(z).$$

Suppose that there exists an (R-two-sided) homomorphic mapping φ of $\mathfrak{M}$ in $\mathfrak{N}$ which maps u_0 on v_0; $u_0^\varphi = v_0$. Put

(6) $$u_h^\varphi = \sum_k v_k x_{kh} \qquad (x_{kh} \in R)$$

Then φ maps $zu_0 = \sum u_h\mu_h(z)$ on $\sum v_k x_{kh}\mu_h(z)$, while $(zu_0)^\varphi = zu_0^\varphi = zv_0 = \sum v_k\nu_k(z)$ too. So $\nu_k(z) = \sum x_{kh}\mu_h(z)$, or

(7) $$\nu_k = \sum_h \mu_h x_{khl}.$$

Thus

(8) $$\nu_1R_l + \nu_2R_l + \ldots + \nu_nR_l \subseteq \mu_1R_l + \mu_2R_l + \ldots + \mu_mR_l.$$

If we consider, firstly, the case that $\mathfrak{M} = \mathfrak{N}$ and φ is the identity mapping, our observation shows that the module

(9) $$\sum_h \mu_hR_l = \mu_1R_l + \mu_2R_l + \ldots + \mu_mR_l$$

does not depend on the special choice of the independent basis $(u_1, u_2, \ldots, u_m)$. We call the module (9) the *relation module* of u_0 in $\mathfrak{M}$.

If we consider secondly the case that $\mathfrak{M} \subseteq \mathfrak{N}$ and φ is again the identity mapping, we find that the relation module of u_0 in a module $\mathfrak{N}$ containing $\mathfrak{M}$ (and having an independent R-right-basis) is contained in that of u_0 in $\mathfrak{M}$. If here $\mathfrak{M}$ is a direct summand in $\mathfrak{N}$ as R-right-module, then the relation moduli of u_0 in $\mathfrak{M}$ and $\mathfrak{N}$ coincide. This last remark, which is rather useful, we see readily by observing that $\mathfrak{N}/\mathfrak{M}$ is regular with integral rank and so $\mathfrak{N}$ has an independent R-right-basis which contains a basis for $\mathfrak{M}$; in fact, every independent R-right-basis of $\mathfrak{M}$ may be extended to one of $\mathfrak{N}$.

Now, if in particular Ru_0 contains an independent R-right-basis of $\mathfrak{M}$, then the m module-endomorphisms $\mu_1, \mu_2, \ldots, \mu_m$ of R are R_l-right-independent, and moreover any independent R_l-right-basis of the relation module is obtained from suitable choice of independent R-right-basis of $\mathfrak{M}$. Let namely

(10) $$u_i = t_iu_0 \qquad (t_i \in R).$$

Then

(11) $$\mu_\lambda(t_i) = \delta_{\lambda i}$$

and t_i is mapped on y_i by $\sum \mu_k y_{ki}$, which implies the right-independence of $\mu_1, \mu_2, \dots, \mu_m$ over R_l.

Further, under the same assumption also the converse of the above relationship between the inclusion (8) and homomorphism is valid. Assuming (8), where μ and ν are given in (4), (5), and also (7), we define φ as R-right-homomorphic mapping of $\mathfrak{M}$ into $\mathfrak{N}$ by virtue of (6). Then

$$u_0^\varphi = (\sum u_\lambda \mu_\lambda(1))^\varphi = \sum v_k x_{k\lambda} \mu_\lambda(1) = \sum v_k \nu_k(1) = v_0.$$

More generally

$$(zu_0)^\varphi = (\sum u_\lambda \mu_\lambda(z))^\varphi = \sum v_k x_{k\lambda} \mu_\lambda(z) = \sum v_k \nu_k(z) = zv_0.$$

Our purpose is to show that φ is also R-left-homomorphic, and we may, for that purpose, assume $u_1, u_2, \dots, u_m \in Ru_0$. On putting $u_\lambda = t_\lambda u_0$ ($t_\lambda \in R$), as in (10), we have

$$u_\lambda^\varphi = (t_\lambda u_0)^\varphi = t_\lambda v_0 = \sum v_k \nu_k(t_\lambda).$$

Comparing this with (6) we obtain

$$x_{k\lambda} = \nu_k(t_\lambda).$$

Therefore

$$\begin{aligned}(zu_\lambda)^\varphi = (zt_\lambda u_0)^\varphi &= (\sum u_i \mu_i(zt_\lambda))^\varphi = \sum v_k x_{ki} \mu_i(zt_\lambda) \\ &= \sum v_k \nu_k(zt_\lambda) = zt_\lambda v_0 = z(\sum v_k \nu_k(t_\lambda)) = z(\sum v_k x_{k\lambda}) = zu_\lambda^\varphi.\end{aligned}$$

This shows that the mapping is R-left-homomorphic, as is desired.

On returning to the case of general u_0 which may not, necessarily, even generate $\mathfrak{M}$, we show further that its relation module $\sum \mu_\lambda R_l$ in $\mathfrak{M}$ is R_l-left-allowable too. For, if we put

(12) $$zu_\lambda = \sum u_j \rho_{j\lambda}(z)$$

then $\sum u_j \mu_j(zy) = zyu_0 = z \sum u_\lambda \mu_\lambda(y) = \sum u_j \rho_{j\lambda}(z) \mu_\lambda(y)$ and

(13) $$z_l \mu_j = \mu_\lambda(\rho_{j\lambda}(z))_l,$$

which proves our assertion.

Here

(14) $$z \to (\rho_{j\lambda}(z))$$

is a self-representation of R, i.e. a (matric) representation of R in R. Consider the relation module of a basis element, say $u_0 = u_i$. Then $\mu_\lambda = \rho_{\lambda i}$, and the relation module $\sum \mu_\lambda R_l$ is nothing but the R_l-right-module generated by $\rho_{1i}, \rho_{2i}, \dots, \rho_{mi}$; this module may thus be called the *i-column module* of the representation (14).

THEOREM 4. *Let $\mathfrak{M}$ be an R-double-module possessing an independent R-right-basis. The relation module $\sum \mu_k R_l$ in* (9) *of an element u_0 in $\mathfrak{M}$, with μ_i given as in* (4), *is independent of the special choice of the independent R-right-basis (u_h) of $\mathfrak{M}$, and is a double-module of R_l. If in particular Ru_0 contains an independent R-right-basis, then $\mu_1, \mu_2, \ldots, \mu_m$ are R_l-right-independent. Let $\mathfrak{N}$ be a second R-double-module with an independent R-right-basis. If φ is an R-two-sided homomorphic mapping of $\mathfrak{M}$ into $\mathfrak{N}$, then the relation module $\sum \nu_k R_l$ of $v_0 = u_0^\varphi$ in $\mathfrak{N}$ is contained in the relation module $\sum \mu_k R_l$ of u_0 in $\mathfrak{M}$* (see (8)). *In particular, if $\mathfrak{M} \subseteq \mathfrak{N}$ then the relation module of u_0 in $\mathfrak{N}$ is contained in that of u_0 in $\mathfrak{M}$. If $\mathfrak{M}$ is direct summand in $\mathfrak{N}$ as R-right-module, then these relation moduli coincide. In case Ru_0 contains an independent R-right-basis of $\mathfrak{M}$ the inclusion $\sum \nu_k R_l \subseteq \sum \mu_k R_l$ is also sufficient in order that there exist an R-two-sided homomorphic mapping of $\mathfrak{M}$ into $\mathfrak{N}$ which maps u_0 upon v_0. Thus the structure of an R-double-module which has an independent R-right-basis contained in Ru_0, with an element u_0 of the module, is uniquely determined by its relation module of the element u_0.*

The last statement means that, if two self-representations of R are defined by R-double-moduli possessing independent R-right-bases contained in the R-left-moduli generated, respectively, by their first, say, basis elements and if their 1-column moduli coincide, then the representations are equivalent (in the usual sense of equivalence of representations).

Further we obtain readily

THEOREM 5. *Let $\mathfrak{M}$, $\mathfrak{N}$ be R-double-moduli, with independent R-right-bases, and let $u_0 \in \mathfrak{M}$, $v_0 \in \mathfrak{N}$. Consider the direct sum $\mathfrak{M} \oplus \mathfrak{N}$ and its element $w_0 = u_0 + v_0$. Then the relation module of w_0 in $\mathfrak{M} \oplus \mathfrak{N}$ is the sum, not necessarily direct, $\sum \mu_k R_l + \sum \nu_k R_l$ of the relation moduli $\sum \mu_k R_l$, $\sum \nu_k R_l$ of u_0, v_0 in $\mathfrak{M}$, $\mathfrak{N}$.*

We next consider the direct product $\mathfrak{M} \times \mathfrak{N} = \mathfrak{M} \times_R \mathfrak{N}$ of $\mathfrak{M}$, $\mathfrak{N}$ and its element $w_0 = u_0 v_0$. We have

THEOREM 6. *The relation module of $w_0 = u_0 v_0$ in $\mathfrak{M} \times \mathfrak{N} = \mathfrak{M} \times_R \mathfrak{N}$ coincides with the product module $(\sum \mu_k R_l)(\sum \nu_k R_l) = \sum \mu_h \nu_k R_l$ of the relation moduli $\sum \mu_k R_l$, $\sum \nu_k R_l$ of u_0, v_0 in $\mathfrak{M}$, $\mathfrak{N}$.*

For, $\mathfrak{M} \times \mathfrak{N} = u_1\mathfrak{N} \oplus u_2\mathfrak{N} \oplus \ldots \oplus u_m\mathfrak{N} = u_1v_1R \oplus u_1v_2R \oplus \ldots \oplus u_mv_nR$, and $u_1v_1, u_1v_2, \ldots, u_mv_n$ are R-right-independent. And

$$zw_0 = zu_0v_0 = \sum u_h\mu_h(z)v_0 = \sum u_hv_k\nu_k(\mu_h(z)).$$

This shows that the relation module of w_0 in $\mathfrak{M} \times \mathfrak{N}$ is really $\sum \mu_h\nu_k R_l$. But this is the product module $(\sum \mu_k R_l)(\sum \nu_k R_l)$, since $\sum \nu_k R_l$ is R_l-left-allowable too.

Now, let S be the totality of elements a in R such that $au_0 = u_0 a$. S is a subring of R. If $a \in S$ then

$$zau_0 = zu_0a = u_1\mu_1(z)a + u_2\mu_2(z)a + \ldots + u_m\mu_m(z)a,$$

whence $\mu_h(za) = \mu_h(z)a$, or $a_r\mu_h = \mu_h a_r$. Thus the relation module $\sum \mu_h R_l$ is contained in $V(S_r)$. If conversely a is an element of R such that μ_h are a_r-endomorphisms, then

$$zau_0 = \sum u_h\mu_h(za) = \sum u_h\mu_h(z)a = zu_0a \qquad (z \in R),$$

in particular $au_0 = u_0a$, and so $a \in S$. Thus

$$(15) \qquad S = \{a \in R;\ au_0 = u_0a\} = \{a \in R;\ \mu_h a_r = a_r\mu_h(h = 1, 2, \ldots, m)\}.$$

Suppose now that Ru_0 contains an independent R-right-basis of $\mathfrak{M}$ and that our relation module $\sum \mu_h R_l$ forms a ring. If $\mathfrak{N} = Rv_0R$ is a second R-double-module which is isomorphic to $\mathfrak{M}$ by $u_0 \leftrightarrow v_0$, then the ring assumption of $\sum \mu_h R_l$ means, by Theorems 4, 6, that $u_0 \to u_0v_0$ gives an (R-two-sided) homomorphic mapping of $\mathfrak{M}$ into the direct product $\mathfrak{M} \times_R \mathfrak{N}$. Let our basis (u_h) of $\mathfrak{M}$ be taken from Ru_0; put $u_h = t_hu_0$ ($t_h \in R$), as in (10). Let (v_h) be the corresponding basis of $\mathfrak{N}$. By our mapping of $\mathfrak{M}$ into $\mathfrak{M} \times \mathfrak{N}$ zu_0 should be mapped upon

$$zu_0v_0 = \sum u_h\mu_h(z)v_0 = \sum u_hv_k\mu_k(\mu_h(z)),$$

while $zu_0 = \sum u_h\mu_h(z) = \sum t_hu_0\mu_h(z)$ and this should be mapped on

$$\sum t_hu_0v_0\mu_h(z) = \sum u_hv_0\mu_h(z) = \sum u_hv_k\mu_k(1)\mu_h(z).$$

We have, since u_hv_k are R-right-independent, $\mu_k(\mu_h(z)) = \mu_k(1)\mu_h(z)$. Then

$$\mu_h(z)u_0 = \sum u_k\mu_k(\mu_h(z)) = \sum u_k\mu_k(1)\mu_h(z) = u_0\mu_h(z),$$

hence $\mu_h(z) \in S$. Let (u'_h) be a second independent R-right-basis of $\mathfrak{M}$ and let μ'_h be the corresponding endomorphisms. Put $u_h = \sum u'_kx_{kh}$. Then $\mu'_h(z) = \sum x_{hk}\mu_k(z)$ and in particular $\mu'_h(t_i) = \sum x_{hk}\delta_{ki} = x_{hi}$. Thus $\mu'_h(R) \subseteq S$ ($h = 1, 2, \ldots, m$) if and only if $x_{hk} \in S$. This last means that $y_{hk} \in S$ for the inverse matrix (y_{hk}) of (x_{hk}). Thus the condition amounts to

$$u'_h \in u_1S \oplus u_2S \oplus \ldots \oplus u_mS = Ru_0S.$$

Here we have, as a matter of fact, $\mu'_h(R) = S$, since the S-right-module generated by x_{hi} ($i = 1, 2, \ldots, m$) certainly exhausts S.

Conversely, if every μ_h maps R in S, then

$$\mu_k(\mu_h(z)) = \mu_k(1\mu_h(z)) = \mu_k(1)\mu_h(z),$$

whence $\mu_h\mu_k \in \mu_hR_l$, and $\sum \mu_hR_l$ forms a ring.

Further, again under the assumption that $\sum \mu_hR_l$ is a ring and $u_h = t_hu_0 \in Ru_0$, we have

$$\sum t_h\mu_h(z)u_0 = \sum t_hu_0\mu_h(z) = \sum u_h\mu_h(z) = zu_0$$

and $z - \sum t_h\mu_h(z)$ is, for every $z \in R$, in the left-ideal $\mathfrak{l} = \{z \in R, zu_0 = 0\}$. Here $t_h(h=1, 2, \ldots, m)$ are S-right-independent mod $\mathfrak{l}$, as we see readily from $\mu_i(t_h) = \delta_{ih}$. Thus (t_h) forms an independent S-right-basis of R mod $\mathfrak{l}$. If

in particular $\mathfrak{l} = 0$, that is, if (the single element) u_0 is R-left-independent, then (t_h) forms an independent S-right-basis of R. Since the R_l-rank m is in that case equal to the S_r-rank of R, our relation module $\sum \mu_h R_l$ must then exhaust the whole Galois ring $V(S_r)$, here we may also argue as in §2 without appealing to the rank relation. So we have

THEOREM 7. *Let $\mathfrak{M} = Ru_0R$ have an independent R-right-basis contained in Ru_0. The relation module of u_0 in $\mathfrak{M}$ forms a ring if*[3] *and only if we may choose such a basis (u_h) so that $\mu_h(R) \subseteq S$ for every h, where S is the subring* (15) *of R; as matter of fact $\mu_h(R) = S$ then. This last is the case, under the ring assumption of the relation module, if and only if $\{u_h\} \subseteq Ru_0S$. Provided that zu_0 $(z \in R)$ vanishes only when $z = 0$, our ring assumption implies also that R has an independent S-right-basis; in fact (t_h) $(h = 1, 2, \ldots, m)$ forms such a basis when (t_hu_0) forms an independent R-right-basis of $\mathfrak{M}$, and moreover the homomorphic mapping $1^l1^r \to u_0$ of the self-product $R \times_S R = R^l \times_S R^r$ of R over S* (§2) *upon $\mathfrak{M}$ becomes an isomorphism. In short, our relation module is a Galois ring if and only if it is a ring and $zu_0 = 0$ implies $z = 0$.*

4. Relationship between relation moduli over R and its subring. Let S be a subring of R and let R possess an independent S-right-basis; $R = x_1S \oplus x_2S \oplus \ldots \oplus x_sS$. Then an R-double-module $\mathfrak{M}$ with an independent R-right-basis $(u_1, u_2, \ldots, u_m)$ is certainly an S-double-module with independent S-right-basis (u_hx_i). Let u_0 be an element of $\mathfrak{M}$ and let its relation module in $\mathfrak{M}$, as R-module, be given by $\sum \mu_h R_l$. We now consider the relation module of u_0 in $\mathfrak{M}$ as S-module. On putting

$$z = x_1\pi_1(z) + x_2\pi_2(z) + \ldots + x_s\pi_s(z) \qquad (\pi_i(z) \in S), \tag{16}$$

we have $zu_0 = \sum u_h\mu_h(z) = \sum_h u\pi_i(\mu_h(z))$. Thus the relation module of u_0 in the S-module $\mathfrak{M}$ is given by

$$\sum_{h,i} \mu_h\pi_i S_l \tag{17}$$

(where μ_h are considered as homomorphisms of S into R).

In case u_0 is one of the basis elements, say u_1, the situation may be described also in terms of representation. Namely, on assuming $x_1 = 1$, without loss of generality, we consider the regular representation $(\lambda_{ij}(z))$ of R in S, with respect to our basis (x_i):

$$zx_j = \sum x_i\lambda_{ij}(z) \qquad (\lambda_{ij}(z) \in S). \tag{18}$$

Denote the self-representation of R defined by our basis $(u_1(= u_0), u_2, \ldots, u_m)$ of $\mathfrak{M}$ by $(\rho_{hk}(z))$, as in (12). The S-(right-)basis (u_hx_i) of $\mathfrak{M}$ defines then the representation

$$(\lambda_{ij}(\rho_{hk}(z))) \tag{19}$$

[3] This "if" part is valid without our assumption of existence of an independent R-right-basis of $\mathfrak{M}$ in Ru_0, or even without assuming $\mathfrak{M} = Ru_0R$.

of degree ms in S. Restricted to S, this gives the self-representation of S defined by the basis $(u_h x_i)$ of S-module $\mathfrak{M}$. Since here $u_1 x_1 = u_0$, our relation module of u_0 in the S-module $\mathfrak{M}$ is obtained as the first, i.e. (1, 1)-, column module of this representation.

THEOREM 8. *The relation module of u_0 in $\mathfrak{M}$ as S-module is given by* (17), *restricted to S, with π_i in* (16). *If in particular $u_0 = u_1$ and $x_1 = 1$, it is also defined as the* (1, 1)-*column module of the self-representation* (19), *restricted to S, of S, where (ρ_{hk}) is the self-representation of R defined by the (R-right-)basis (u_h) of $\mathfrak{M}$ and (λ_{ij}) is the regular representation of R in S defined by the (S-right-)basis (x_i).*

We supplement the theorem with the following observation: Let $\mathfrak{m}$ be an S-double-module with independent S-right-basis. Then there always exists an R-double-module $\mathfrak{M}$ with independent R-right-basis, which contains, as S-double-module, $\mathfrak{m}$, and which contains $\mathfrak{m}$ as S-right-module indeed as direct summand. (Then the relation module of u_0 $(\in \mathfrak{m})$ in $\mathfrak{m}$ coincides with that in $\mathfrak{M}$, as S-module. Therefore it is thus obtained from the relation module of u_0 in R-module $\mathfrak{M}$ by virtue of the above procedure of referring to S in terms of π_i(in (16)).

Let $(v_1, v_2, \ldots, v_n)$ be an independent S-right-basis of $\mathfrak{m}$,

$$\mathfrak{m} = v_1 S \oplus v_2 S \oplus \ldots \oplus v_n S.$$

$\mathfrak{m} \times_S R$ is an R-right-module $v_1 R \oplus v_2 R \oplus \ldots \oplus v_n R$ with $v_1, v_2, \ldots, v_n$ right-independent over R. Therefore

$$R \times_S \mathfrak{m} \times_S R = x_1(\mathfrak{m} \times_S R) \oplus x_2(\mathfrak{m} \times_S R) \oplus \ldots \oplus x_s(\mathfrak{m} \times_S R)$$

is an R-double-module with independent R-right-basis $(x_i v_k)$. On assuming $x_1 = 1$, it follows that the S-two-sided submodule $x_1 \mathfrak{m} = \mathfrak{m}$ is its direct summand as S-right-module.

5. Supplementary remarks. If R is a primary-decomposable ring, then a regular R-right-module is always a direct summand in a second regular R-right-module which contains it. If R is a simple ring then a (finite) R-right-module is always regular. These remarks are significant in connection with the theorems in §3, in particular with Theorems 4, 6. If, moreover, R is a quasifield, then any R-right-module certainly has an independent basis and any subring is of course also a quasifield. In dealing with relation moduli over a quasifield R we may thus always restrict ourselves to principal R-double-moduli which possess R-right-bases contained in the R-left-module generated by the element in question. Furthermore, the hypergroup formulation of our Galois theory can then be given under a certain assumption.

On the other hand, it may be of some use, in view of the usual Galois theory, to observe the case in which each $u_h R$, with a basis element u_h of $\mathfrak{M}$, is R-left-allowable too. Let an R-double-module $\mathfrak{M}$ possess such an independent R-right-basis $(u_1, u_2, \ldots, u_m)$ and let u_0 be the sum $u_0 = u_1 + u_2 + \ldots + u_m$.

The relation module of u_0 in $\mathfrak{M}$ is $\sum \mu_h R_l$, where we put $zu_0 = \sum u_h \mu_h(z)$. Since $Ru_h \subseteq u_h R$, we have

$$zu_h = u_h \mu_h(z)$$

and each μ_h is simply the self-representation of degree 1, i.e. (ring-) endomorphism, of R defined by the representation module $u_h R = Ru_h R$ (with respect to the basis element u_h). If $\mathfrak{M}$ has an independent R-right-basis contained in Ru_0 then these m endomorphisms μ_h of R are right-independent over R_l.

In this context we note some sufficient conditions that certain given (ring-) endomorphisms, say $\nu_1, \nu_2, \ldots, \nu_n$, of R be right-independent over R_l. Let (R, ν), with a (ring-)endomorphism ν of R, denote an R-double-module which coincides with R itself as R-left-module and on which right-operation of $z \in R$ is defined by $x\ (\in R) \rightarrow x.z = x\nu(z)$. Thus (R, ν) may also be looked upon as a module Rw with R-left-independent element w such as $wz = \nu(z)w$. Now we have, firstly: *If each $\nu_i(R)$ possesses no non-zero left-annihilator in R and if*

(*) *the R-double-moduli (R, ν_i), (R, ν_j) with distinct i, j have non-zero (R-two-sided)) isomorphic submoduli, then $\nu_1, \nu_2, \ldots, \nu_n$ are R-right-independent.*

For, from our assumptions we deduce that $\nu_i x_l = 0$ implies $x = 0$, for each i, and the R_l-double-moduli $\nu_i R_l$ and $\nu_j R_l$ with $i \neq j$ have no (R_l-two-sided) isomorphic non-zero submoduli. The sum $\sum \nu_i R_l$ is then necessarily direct [8, §3, Remark 6].

Secondly, if ν_i are (ring-)automorphisms and $\{\nu_i\}$ forms a group which induces a Galois group of the residue-ring R/N of R modulo its radical N in the sense of [8] (that is, a similar assumption (**) obtained from (*) by replacing "submoduli" by "residue-submoduli" is satisfied), then again ν_i are R_l-right-independent [8, Lemma 4 and Remark 5 concerning it].

A similar construction can be used to show that a certain R-double-module is principal. Interchanging "left" and "right", in order to be in accord with our situation, we consider n elements $v_1, v_2, \ldots, v_n$ which are right-independent over R and satisfy $zv_i = v_i\nu_i(z)$, with (ring-)endomorphisms ν_i of R. Suppose that the (left-right-)symmetric counterpart of (**), mentioned above, is satisfied. Then if v_0 is an element in the (direct) sum $\mathfrak{N} = \sum \nu_i R$ of a form $v_0 = v_1 z_1 + v_2 z_2 + \ldots + v_n z_n$ with regular elements z_i of R, we have

$$\mathfrak{N} = Rv_0R.$$

For, under our assumption, Rv_0R exhausts the whole $\mathfrak{N}$ mod $\sum \nu_i N$ firstly, where N denotes the radical of R, and then actually, since $\sum v_i N$ is (whence is contained in) the intersection of all maximal R-right-submoduli of $\mathfrak{N}$.

Of course all these assumptions are covered by the assumption that $G = \{\nu_i\}$ forms a Galois group of R, under which in [8] the complete correspondence of between-rings, over which R has independent right-basis, with subgroups of G (not only with certain subrings of $\sum \nu_i R_l$) was established.

Let S be a subring of R such that R has not only an independent S-right-basis

of s terms but also an independent S-left-basis of the same number s of terms. Suppose further that $\mathbf{B} = V(S_r)$ contains an independent R_l-right-basis $(\beta_1, \beta_2, \ldots, \beta_s)$ (i.e. a Galois system belonging to S) which forms also an independent R_l-left-basis of $\mathbf{B}$, and that $\beta_1 S_l \oplus \beta_2 S_l \oplus \ldots \oplus \beta_s S_l$ forms a ring $(\ni 1)$ and moreover it equals $S_l\beta_1 \oplus S_l\beta_2 \oplus \ldots \oplus S_l\beta_s$. Then there exists an element x in R such that $\beta_1(x), \beta_2(x), \ldots, \beta_s(x)$ form an independent S-left-basis of R.

To show this, we observe that R is B-regular with rank $1/s$, or, what is the same, the direct sum R^s of s copies of R is $\mathbf{B}$-isomorphic to the $\mathbf{B}$-right-module $\mathbf{B}$. Hence naturally R^s is $(S_l\beta_1 \oplus S_l\beta_2 \oplus \ldots \oplus S_l\beta_s)$-isomorphic to $\mathbf{B}$. On the other hand

$$\begin{aligned}\mathbf{B} = R_l\beta_1 \oplus R_l\beta_2 \oplus \ldots \oplus R_l\beta_s &= y_{1l}(S_l\beta_1 \oplus \ldots \oplus S_l\beta_s) \oplus \ldots \\ &\oplus y_{sl}(S_l\beta_1 \oplus \ldots \oplus S_l\beta_s),\end{aligned}$$

where $(y_1, y_2, \ldots, y_s)$ is an independent S-left-basis of R. Hence $\mathbf{B}$ is a regular $(S_l\beta_1 \oplus \ldots \oplus S_l\beta_s)$-right-module of rank s. It follows that R is $(S_l\beta_1 \oplus \ldots \oplus S_l\beta_s)$-(right-)isomorphic to $S_{ll}\beta_1 \oplus \ldots \oplus S_l\beta_s = \beta_1 S_l \oplus \ldots \oplus \beta_s S_l$. Let x be the element of R which is mapped on the unit element of $\beta_1 S_l \oplus \ldots \oplus \beta_s S_l$ in such an isomorphism. Then $(x^{\beta_1}, x^{\beta_2}, \ldots, x_1^{\beta_s})$ forms an independent S_l-right-basis, that is, S-left-basis of R. This statement, though complicated, may be regarded as a generalization of the theorem of normal basis.

If here β_h are (ring-)automorphisms of R, then $\beta_h R_l = R_l\beta_h$ and moreover each β_h is elementwise commutative with S_l. Hence the left-symmetric half of the assumption concerning β_h follows automatically.

References

[1] G. Azumaya, *Galois theory of uni-serial rings*, J. Math. Soc. Japan, vol. 1 (1949).

[2] N. Jacobson, *The fundamental theorem of Galois theory for quasifields*, Ann. Math., vol. 41 (1940).

[3] ———, *An extension of Galois theory to non-normal and non-separable fields*, Amer. J. Math., vol. 66 (1944).

[4] ———, *Relations between the composites of a field and those of a subfield*, Amer. J. Math., vol. 66 (1944).

[5] ———, *Galois theory of purely inseparable fields of exponent one*, Amer. J. Math., vol. 66 (1944).

[6] ———, *Note on division rings*, Amer. J. Math., vol. 69 (1947).

[7] T. Nakayama, *Semilinear normal basis for quasifields*, Amer. J. Math., vol. 71 (1949).

[8] ———, *Galois theory for general rings with minimum condition*, J. Math. Soc. Japan, vol. 1 (1949).

[9] ———, *Commuter systems in a ring with radical*, Duke Math. J., vol. 16 (1949).

[10] ———, *Generalized Galois theory for rings with minimum condition*, in Amer. J. Math.

[11] T. Nakayama and G. Azumaya, *On irreducible rings*, Ann. Math., vol. 48 (1947).

Nagoya University

ON COUNTABLY PARACOMPACT SPACES

C. H. DOWKER

LET X be a topological space, that is, a space with open sets such that the union of any collection of open sets is open and the intersection of any finite number of open sets is open. A covering of X is a collection of open sets whose union is X. The covering is called countable if it consists of a countable collection of open sets or finite if it consists of a finite collection of open sets; it is called locally finite if every point of X is contained in some open set which meets only a finite number of sets of the covering. A covering $\mathfrak{V}$ is called a refinement of a covering $\mathfrak{U}$ if every open set of $\mathfrak{V}$ is contained in some open set of $\mathfrak{U}$. The space X is called countably paracompact if every countable covering has a locally finite refinement.

The purpose of this paper is to study the properties of countably paracompact spaces. The justification of the new concept is contained in Theorem 4 below, where it is shown that, for normal spaces, countable paracompactness is equivalent to two other properties of known topological importance.

1. A space X is called compact if every covering has a finite refinement, paracompact if every covering has a locally finite refinement, and countably compact if every countable covering has a finite refinement. It is clear that every compact, paracompact or countably compact space is countably paracompact. Just as one shows[1] that every closed subset of a compact [paracompact, countably compact] space is compact [paracompact, countably compact], so one can show that every closed subset of a countably paracompact space is countably paracompact. It is known that the topological product of two compact spaces is compact and the topological product of a compact space and a paracompact space is paracompact [2, Theorem 5]. The following is an analogous theorem.

THEOREM 1. *The topological product $X \times Y$ of a countably paracompact space X and a compact space Y is countably paracompact.*

Proof. Let $\{U_i\}$ $(i = 1, 2, \ldots)$ be a countable covering of $X \times Y$. Let V_i be the set of all points x of X such that $x \times Y \subset \bigcup_{j<i} U_j$. If $x \in V_i$ every point (x, y) of $x \times Y$ has a neighbourhood $N \times M$, (N open in X, M open in Y), which is contained in the open set $\bigcup_{j<i} U_j$. A finite number of these open sets M cover Y; let N_x be the intersection of the corresponding finite number of sets N. Then $x \in N_x$, N_x is open and $N_x \times Y \subset \bigcup_{j\leq i} U_j$; and hence $N_x \subset V_i$. Therefore V_i is open. Also, for any $x \in X$, since $x \times Y$ is compact, $x \times Y$ is

Received December 19, 1949.

[1]See [1] page 86, Satz IV and [2] Theorem 2.

contained in some finite number of sets of the covering $\{U_i\}$; hence x is in some V_i. Therefore $\{V_i\}$ is a covering of X.

Since $\{V_i\}$ is countable and X is countably paracompact, $\{V_i\}$ has a locally finite refinement $\mathfrak{W}$. For each open set W of $\mathfrak{W}$ let $g(W)$ be the first V_i containing W and let G_i be the union of all W for which $g(W) = V_i$. Then G_i is open, $G_i \subset V_i$ and $\{G_i\}$ is a locally finite covering of X.

If $j \leqslant i$, let $G_{ij} = (G_i \times Y) \cap U_j$; then G_{ij} is an open set in $X \times Y$. If (x, y) is any point of (X, Y) then, for some i, $x \in G_i$ and hence $(x, y) \in G_i \times Y$. Also, since $x \in G_i \subset V_i$, $(x, y) \in x \times Y \subset \bigcup_{j \leqslant i} U_j$, and hence, for some $j \leqslant i$, $(x, y) \in U_j$. Hence $(x, y) \in G_{ij}$. Therefore $\{G_{ij}\}$ is a covering of $X \times Y$. Since $G_{ij} \subset U_j$, $\{G_{ij}\}$ is a refinement of $\{U_i\}$. Also, if $(x, y) \in X \times Y$, x is in an open set $H(x)$ which meets only a finite number of the sets of $\{G_i\}$. Then $H(x) \times Y$ is an open set containing (x, y) which can meet G_{ij} only if $H(x)$ meets G_i. But for each i there is only a finite number of sets G_{ij}. Hence $H(x) \times Y$ meets only a finite number of sets of $\{G_{ij}\}$; hence $\{G_{ij}\}$ is locally finite. Therefore $X \times Y$ is countably paracompact. This completes the proof.

It can similarly be shown that the topological product of a compact space and a countably compact space is countably compact.

2. A topological space X is called normal if for every pair of disjoint closed sets A and B of X there is a pair of disjoint open sets U and V with $A \subset U$ and $B \subset V$ (or, equivalently, there is an open set U with $A \subset U, \overline{U} \subset X - B$).

THEOREM 2. *The following properties of a normal space X are equivalent*:

(a) *The space X is countably paracompact.*

(b) *Every countable covering of X has a point-finite*[2] *refinement.*

(c) *Every countable covering $\{U_i\}$ has a refinement $\{V_i\}$ with $\overline{V}_i \subset U_i$.*

(d) *Given a decreasing sequence $\{F_i\}$ of closed sets with vacuous intersection, there is a sequence $\{G_i\}$ of open sets with vacuous intersection such that $F_i \subset G_i$.*

(e) *Given a decreasing sequence $\{F_i\}$ of closed sets with vacuous intersection, there is a sequence $\{A_i\}$ of closed G_δ-sets*[3] *with vacuous intersection such that $F_i \subset A_i$.*

Proof. (a) → (b). A locally finite covering is *a fortiori* point-finite.

(b) → (c). Let $\{U_i\}$ be any countable covering of X. Then, by (b), $\{U_i\}$ has a point-finite refinement $\mathfrak{W}$. For each open set W of $\mathfrak{W}$ let $g(W)$ be the first U_i containing W, and let G_i be the union of all W such that $g(W) = U_i$. Then $\{G_i\}$ is a point-finite covering of X and $G_i \subset U_i$. It is known [3, p. 26, (33-4); 2, Theorem 6] that every point-finite covering $\{G_i\}$ (whether countable or not) of a normal space X has a refinement $\{V_i\}$ with the closure of each V_i contained in the corresponding G_i. Then $\overline{V}_i \subset G_i \subset U_i$, hence $\overline{V}_i \subset U_i$.

---

[2] A covering of X is called point-finite if each point of X is in only a finite number of sets of the covering.

[3] A set A is called a G_δ-set if it is the intersection of some countable collection of open sets.

(c) → (d). Let $\{F_i\}$ be a sequence of closed sets with $F_{i+1} \subset F_i$ and $\bigcap_i F_i = 0$. Then, if $U_i = X - F_i$, $\{U_i\}$ is a covering of X. Then, by c, there is a covering $\{V_i\}$ with $\overline{V}_i \subset U_i$. Let G_i be the open set $X - \overline{V}_i$. Then, since $\overline{V}_i \subset U_i$, $F_i \subset G_i$ and, since $\bigcup \overline{V}_i = X$, $\bigcap G_i = 0$.

(d) → (e). Let $\{F_i\}$ be a sequence of closed sets with $F_{i+1} \subset F_i$ and $\bigcap F_i = 0$. Then, by d, there is a sequence $\{G_i\}$ of open sets with $F_i \subset G_i$ and $\bigcap G_i = 0$. Then, by Urysohn's lemma, there is a continuous function ϕ_i, $0 \leqslant \phi_i(x) \leqslant 1$, such that, if $x \in F_i$, $\phi_i(x) = 0$ and, if x non $\in G_i$, $\phi_i(x) = 1$. Let $G_{ij} = \{x \mid \phi_i(x) < 1/j\}$, and let $A_i = \bigcap_j G_{ij} = \{x \mid \phi_i(x) = 0\}$. Then G_{ij} is open, A_i is a closed G_δ-set, $F_i \subset A_i \subset G_i$ and $\bigcap A_i \subset \bigcap G_i = 0$.

(e) → (a). Let $\{U_i\}$ be a countable covering of X and let $F_i = X - \bigcup_{k<i} U_k$. Then F_i is closed, $F_{i+1} \subset F_i$ and, since $\bigcup U_i = X$, $\bigcap F_i = 0$. Then, by (e), there is a sequence $\{A_i\}$ of closed G_δ-sets with $F_j \subset A_j$ and $\bigcap A_j = 0$. Then $X - A_j$ is an F_σ-set; let $X - A_j = \bigcup_i B_{ji}$ where each B_{ji} is closed. Since X is normal we may assume that B_{ji} is contained in the interior of $B_{j,\,i+1}$. Let H_{ji} be the interior of B_{ji}; then $H_{ji} \subset B_{ji} \subset H_{j,\,i+1}$ and $X - A_j = \bigcup_i H_{ji}$. And $B_{ji} \subset X - A_j \subset X - F_j = \bigcup_{k \leqslant j} U_k$.

Let $V_i = U_i - \bigcup_{j<i} B_{ji}$; then V_i is open. If $j < i$, $B_{ji} \subset \bigcup_{k \leqslant j} U_k \subset \bigcup_{k<i} U_k$; hence $\bigcup_{j<i} B_{ji} \subset \bigcup_{k<i} U_k$. Hence $V_i \supset U_i - \bigcup_{k<i} U_k$. Thus, since each point x of X is in a first U_i, it is in the corresponding V_i. Therefore $\{V_i\}$ is a covering of X. Clearly $\{V_i\}$ is a refinement of $\{U_i\}$.

For each x of X there is some A_j such that x non ϵ A_j; hence, for some k, $x \in H_{jk}$. Then, if $i > j$ and $i > k$, $H_{jk} \subset B_{ji}$ and hence $H_{jk} \cap V_i = 0$. Thus the open set H_{jk} contains x and meets only a finite number of the sets V_i. Hence $\{V_i\}$ is locally finite. Therefore X is countably paracompact.

Corollary. *Every perfectly normal space is countably paracompact.*

Proof. A perfectly normal space is a normal space in which every closed set is a G_δ-set. Hence condition (e) is trivally satisfied with $A_i = F_i$.

Not every normal space is countably paracompact as the following example shows. Let X be a space whose points x are the real numbers. Let the open sets of X be the null set, the whole space X and the subsets $G_a = \{x \mid x < a\}$ for all real a. Then X is trivially normal since there are no non-empty disjoint closed sets. But the countable covering $\{G_i\}$ $(i = 1, 2, \ldots)$ where $G_i = \{x \mid x < i\}$, has no locally finite refinement. Hence X is not countably paracompact.[4]

3. We give here a sufficient condition for the normality of a product space.

Lemma 3. *The topological product $X \times Y$ of a countably paracompact normal space X and a compact metric space Y is normal.*

Proof. Let A and B be two disjoint closed sets of $X \times Y$. Let $\{G_i\}$ be a

[4]This space is not a Hausdorff space. It would be interesting to have an example of a normal Hausdorff space which is not countably paracompact.

countable base for the open sets of Y and, if γ is any finite set of positive integers, let $H_\gamma = \bigcup_{i \in \gamma} G_i$. For each $x \in X$ let A_x be the closed set of Y defined by $x \times A_x = (x \times Y) \cap A$; similarly let $x \times B_x = (x \times Y) \cap B$. Let

$$U_\gamma = \{x \mid A_x \subset H_\gamma \subset \overline{H}_\gamma \subset Y - B_x\}.$$

Let x_0 be a point of X for which $A_{x_0} \subset H_\gamma$. Then, for each $y \in Y - H_\gamma$, (x_0, y) non$\in A$ and, since A is closed, there is a neighbourhood $N \times M$ of (x_0, y) which does not meet A. A finite number of the open sets M cover the compact set $Y - H_\gamma$. If N_{x_0} is the intersection of the corresponding finite number of open sets N, $N_{x_0} \times (Y - H_\gamma)$ does not meet A. Hence, if $x \in N_{x_0}$, $A_x \subset H_\gamma$. Thus $\{x \mid A_x \subset H_\gamma\}$ is an open set. Similarly $\{x \mid \overline{H}_\gamma \subset Y - B_x\}$ is open and U_γ, which is the intersection of these two open sets, is also open.

Let $x \in X$; then for each point y of A_x there is an open set G_i of the base such that $y \in G_i$ and $\overline{G}_i \cap B_x = 0$. A finite number of these sets G_i cover A_x, i.e., for some finite set γ of positive integers, $A_x \subset \bigcup_{i \in \gamma} G_i = H_\gamma$ and $\overline{H}_\gamma = \bigcup_{i \in \gamma} \overline{G}_i \subset Y - B_x$. Hence $x \in U_\gamma$. Thus the open sets U_γ cover X. Since there are only a countable number of finite subsets γ of positive integers, the covering $\{U_\gamma\}$ of X is countable.

Since X is countably paracompact there is a locally finite covering $\{W_\gamma\}$ of X with $W_\gamma \subset U_\gamma$ and, by condition c of Theorem 2, $\{W_\gamma\}$ has a refinement $\{V_\gamma\}$ (still locally finite) such that $\overline{V}_\gamma \subset W_\gamma$. Let U be the open set $\bigcup_\gamma (V_\gamma \times H_\gamma)$. For any point (x, y) of A and for some V_γ, $x \in V_\gamma \subset U_\gamma$. Then $y \in A_x \subset H_\gamma$ and hence $(x, y) \in V_\gamma \times H_\gamma$; therefore $A \subset U$. Since $\{V_\gamma\}$ is locally finite, each point x of X is contained in an open set $G(x)$ which meets only a finite number of sets V_γ; and hence the neighbourhood $G(x) \times Y$ of (x, y) meets only a finite number of the sets $V_\gamma \times H_\gamma$. It follows that (x, y) is in the closure of U if and only if it is in the closure of some $V_\gamma \times H_\gamma$, i.e., $\overline{U} = \bigcup(\overline{V_\gamma \times H_\gamma})$. But $\overline{V_\gamma \times H_\gamma} = \overline{V}_\gamma \times \overline{H}_\gamma$. Hence $\overline{U} = \bigcup(\overline{V}_\gamma \times \overline{H}_\gamma) \subset \bigcup(U_\gamma \times \overline{H}_\gamma)$. But $(U_\gamma \times \overline{H}_\gamma) \cap B = 0$; hence $\overline{U} \cap B = 0$. Thus the open set U contains A and its closure does not meet B. Hence $X \times Y$ is normal.

4. In Theorem 4 below we extend some results of J. Dieudonné [2]. He showed[5] that paracompactness of a Hausdorff space X implies condition β (see below) on semicontinuous functions on X and our proof that $\alpha \rightarrow \beta$ is a trivial modification of his proof. It also follows immediately from Dieudonné's results that if X is a paracompact Hausdorff space, $X \times I$ is a paracompact Hausdorff space and hence is normal. However, in terms of countable paracompactness we are able to give a necessary and sufficient condition for β and γ to hold. The equivalence of conditions β and γ was conjectured by S. Eilenberg.

THEOREM 4. *The following three properties of a topological space X are equivalent.*

(α). *The space X is countably paracompact and normal.*

[5] See [2], Theorem 9.

(β). *If g is a lower semicontinuous real function on X and h is an upper semicontinuous real function on X and if $h(x) < g(x)$ for all $x \in X$, then there exists a continuous real function f such that $h(x) < f(x) < g(x)$ for all $x \in X$.*

(γ). *The topological product $X \times I$ of X with the closed line interval $I =$ [0, 1] is normal.*

Proof. (α) $\rightarrow$ (β). Let X be a countably paracompact normal space and let g and h be lower and upper semicontinuous functions respectively with $h(x) < g(x)$. If r is a rational number let $G_r = \{x \mid h(x) < r < g(x)\}$. Since g is lower semicontinuous, $\{x \mid g(x) > r\}$ is open, and, since h is upper semicontinuous, $\{x \mid h(x) < r\}$ is open. Hence G_r is open. Since, for every x, $h(x) < g(x)$ there is some rational number $r(x)$ with $h(x) < r(x) < g(x)$; hence $x \in G_{r(x)}$. Thus $\{G_r\}$ is a covering of X. And, since the rational numbers are countable, $\{G_r\}$ is a countable covering. Hence, since X is countably paracompact and normal, there is a locally finite covering $\{U_r\}$ of X with $U_r \subset G_r$ and there is a (locally finite) covering $\{V_r\}$ with $\overline{V}_r \subset U_r$.

There is a continuous function f_r with $-\infty \leqslant f_r(x) \leqslant r$ such that $f_r(x) = -\infty$ if x non $\in U_r$ and $f_r(x) = r$ if $x \in \overline{V}_r$. Let $f(x)$ be the least upper bound of $f_r(x)$ for all r. Each point x_0 of X is contained in an open set $N(x_0)$ which meets only a finite number of the sets U_r. Hence, in $N(x_0)$, for all but a finite number of values of r, $f_r(x) = -\infty$. Thus, in each neighbourhood $N(x_0)$, $f(x)$ is the least upper bound of a finite number of continuous functions, hence f is continuous. In U_r, $f_r(x) \leqslant r < g(x)$ and, in $X - U_r$, $f_r(x) = -\infty < g(x)$. Thus $f_r(x) < g(x)$ and, for each x, $f(x)$ is the least upper bound of a finite number of $f_r(x)$ each less than $g(x)$. Therefore $f(x) < g(x)$. Each x is in some V_r and, for this r, $f_r(x) = r$; hence $f(x) \geqslant f_r(x) = r > h(x)$. Hence $f(x) > h(x)$. Therefore $h(x) < f(x) < g(x)$.

(β) $\rightarrow$ (α). Let X be a space satisfying condition (β) and let A and B be two disjoint closed sets in X. Let h be the characteristic function of A, i.e., $h(x) = 1$ if $x \in A$ and $h(x) = 0$ if x non $\in A$. Let g be defined by $g(x) = 1$ if $x \in B$ and $g(x) = 2$ if x non $\in B$. Then g is lower semicontinuous, h is upper semicontinuous and $h(x) < g(x)$ for all $x \in X$. Hence there is a continuous function f with $h(x) < f(x) < g(x)$. Let $U = \{x \mid f(x) > 1\}$ and $V = \{x \mid f(x) < 1\}$. Then U and V are disjoint open sets and $A \subset U$ and $B \subset V$. Hence X is normal.

Let $\{F_i\}$ $(i = 1, 2, \ldots)$ be a decreasing sequence of closed sets with $\cap F_i = 0$. Let g be defined by $g(x) = 1/(i + 1)$ for $x \in F_i - F_{i+1}$ $(i = 0, 1, \ldots)$, where F_0 means the whole space X. Let $h(x) = 0$ for all $x \in X$. Then g is lower semicontinuous, h is upper semicontinuous and $h(x) < g(x)$ for all x. Hence there is a continuous function f with $0 < f(x) < g(x)$. Let $G_i = \{x \mid f(x) < 1/(i + 1)\}$. Then G_i is open, $F_i \subset G_i$ and, since $f(x) > 0$ for all x, $\bigcap G_i = 0$. Thus condition d of Theorem 2 is satisfied and therefore X is countably paracompact.

(α) $\rightarrow$ (γ). This follows immediately from Lemma 3 and the fact that the interval I is a compact metric space.

$(\gamma) \to (\alpha)$. Let X be a space for which $X \times I$ is normal. Then X is homeomorphic to the closed subset $X \times 0$ of the normal space $X \times I$; therefore X is normal.

Let $\{F_i\}(i=1, 2, \ldots)$, be a decreasing sequence of closed sets with $\bigcap F_i = 0$. Then, since the half open interval $[0, 1/i[$ is open in $I = [0, 1]$, $W_i = (X - F_i) \times [0, 1/i[$ is open in $X \times I$. Let A be the closed set $X \times I - \bigcup_i W_i$. If $x \in X$, then, for some i, $x \in X - F_i$ and $(x, 0) \in W_i$ and hence $(x, 0)$ non$\in A$. Hence, if $B = X \times 0$, A and B are disjoint closed sets of the normal space $X \times I$. Therefore there are disjoint open sets U and V with $A \subset U$ and $B \subset V$. Let $G_i = \{x \mid (x, 1/i) \in U\}$; then G_i is open. For each $x \in X$, $(x, 0) \in B$ and hence, for sufficiently large i, $(x, 1/i) \in V$ and hence non$\in U$. Therefore $\bigcap G_i = 0$. Let $x \in F_i$. Then, if $j \leqslant i$, $F_i \subset F_j$ and x non$\in X - F_j$, and, if $j \geqslant i$, $1/i$ non $\in [0, 1/j[$. Hence $(x, 1/i)$ non$\in \bigcup_j W_j$; hence $(x, 1/i) \in A \subset U$ and hence $x \in G_i$. Therefore $F_i \subset G_i$. Thus condition (d) of Theorem 2 is satisfied and therefore X is countably paracompact. This completes the proof of the theorem.

References

[1] P. Alexandroff and H. Hopf, *Topologie* I (Berlin, 1935).

[2] J. Dieudonné, *Une généralisation des espaces compacts*, Journal de Mathématiques Pures et Appliquées, vol. 23 (1944), 65-76.

[3] S. Lefschetz, *Algebraic Topology* (New York, 1942).

[4] R. H. Sorgenfrey, *On the topological product of paracompact spaces*. Bull. Amer. Math. Soc., vol. 53 (1947), 631-632.

[5] A. H. Stone, *Paracompactness and product spaces*, Bull. Amer. Math. Soc., vol. 54 (1948), 977-982.

Harvard University

THE REPRESENTATIONS OF GL(3, q), GL(4, q), PGL(3, q), AND PGL(4, q)

ROBERT STEINBERG

1. Introduction. This paper is a result of an investigation into general methods of determining the irreducible characters of GL(n, q), the group of all non-singular linear substitutions with marks in GF(q), and of the related groups, SL(n, q), PGL(n, q), PSL(n, q), the corresponding group of determinant unity, projective group, projective group of determinant unity, respectively. This investigation is not complete, but the general problem was answered partially in [9]. In [3], [7], [6], [1], Frobenius, Schur, Jordan, and Brinkmann gave the characters of PSL(2, p); SL(2, q), GL(2, q); SL(2, q), GL(2, q); PSL(3, q), respectively. In this paper in §2 and §3, the characters of GL(2, q) and GL(3, q) are determined, and, from them, those of PGL(2, q) and PGL(3, q) deduced. In §4, an outline of the determination of the characters of GL(4, q) is given together with the degrees and frequencies of the characters of GL(4, q) and PGL(4, q) and a table of the rational characters of GL(4, q).

The simple properties of the underlying geometry, PG($n-1$, q), of which PGL(n, q) is the collineation group, are used throughout the work. The most powerful and frequent tool used in the determination of the characters is the Frobenius method[1] of induced representations [5] which enables one to construct a representation of a group if a representation of a subgroup is known. The explicit formula for the character in this case is $\chi(G) = \dfrac{m}{g_G} \Sigma\psi(G')$, where m is the index of the subgroup, g_G is the number of elements of the group similar to G, ψ is the character of the subgroup, and the summation is made over all elements G' which are similar to G and lie in the subgroup. Of fundamental use in the application of this method are the $q-1$ linear characters of GL(n, q) which correspond to the powers of the determinants of the matrices which define the elements of GL(n, q). Also very useful are pseudo-characters—linear combinations of irreducible characters with negative coefficients permissible—and the fact that a pseudo-character, $\chi(G)$, is an irreducible character if and only if $\Sigma|\chi(G)|^2 = g$ and $\chi(E) > 0$, where E is the unit element of the group.

The descent from the characters of GL(n, q) to those of PGL(n, q) is immediate because of the following two theorems due to Frobenius [4], [5]:

If $\mathfrak{H}$ is a normal subgroup of a group $\mathfrak{G}$, then every character of $\mathfrak{G}/\mathfrak{H}$ is also a character of $\mathfrak{G}$.

Received March 13, 1950.

This paper is part of a Ph.D. thesis written at the University of Toronto under the direction of Professor Richard Brauer.

[1] See [8] for a complete account of the properties of group characters used here.

In order that a character of $\mathfrak{G}$ may belong to the group $\mathfrak{G}/\mathfrak{H}$, it is necessary and sufficient that it have the same value for all elements of $\mathfrak{H}$. Then, it has also equal values for every two elements of $\mathfrak{G}$ which are equivalent mod $\mathfrak{H}$.

In our case, $\mathfrak{G}$ is the group $GL(n, q)$, $\mathfrak{H}$ is the cyclic group of the $q-1$ scalar matrices, and $\mathfrak{G}/\mathfrak{H}$ is the group $PGL(n, q)$. For this reason, and also because the group $GL(n, q)$ is easier to handle, its characters are first determined and then those of $PGL(n, q)$ obtained from them.

In what follows, $\chi_q^{(r)}$ for example, will denote a character of degree q, the superscript being used to distinguish between two characters of the same degree. $GL(1, 2; q)$ denotes the subgroup $\begin{pmatrix} A_1 & 0 \\ * & A_2 \end{pmatrix}$ of $GL(3, q)$; $\rho, \sigma, \tau, \omega$ are primitive elements of $GF(q)$, $GF(q^2)$, $GF(q^3)$, $GF(q^4)$ respectively, such that $\rho = \sigma^{q+1} = \tau^{q^2+q+1} = \omega^{q^3+q^2+q+1}$ and $\sigma = \tau^{q^2+1}$.

2. The characters of $GL(2, q)$ and $PGL(2, q)$. The group $GL(2, q)$ is of order $q(q-1)^2(q+1)$ and each of its elements is similar to a matrix of one of the following four types [2]:

$$A_1: \begin{pmatrix} \rho^a & \\ & \rho^a \end{pmatrix},\ A_2: \begin{pmatrix} \rho^a & \\ 1 & \rho^a \end{pmatrix},\ A_3: \begin{pmatrix} \rho^a & \\ & \rho^b \end{pmatrix}_{a \neq b},\ B_1: \begin{pmatrix} \sigma^a & \\ & \sigma^{aq} \end{pmatrix}_{a \neq \text{mult. } (q+1)}.$$

The number of classes of each type and the number of elements in each class is given by Table I. The total number of classes is $(q-1)(q+1) = k$.

TABLE I

Element	Number of classes	Number of elements in each class
A_1	$q-1$	1
A_2	$q-1$	$(q-1)(q+1)$
A_3	$\frac{1}{2}(q-1)(q-2)$	$q(q+1)$
B_1	$\frac{1}{2}q(q-1)$	$q(q-1)$

Now, if we consider each matrix as a linear transformation of $PG(1, q)$, we get a representation of degree $q+1$ representing the permutation of the points of $PG(1, q)$. The character of any element of $GL(2, q)$ is just the number of points left fixed by it. This permutation group is doubly transitive and hence splits into the unit representation and an irreducible representation [9] of degree q. Multiplication of each of these characters by each of the $q-1$ linear characters given by the powers of the determinants gives us $q-1$ irreducible characters of degree 1 and $q-1$ of degree q. (See Table I.)

We next consider the subgroup $GL(1, 1; q) = \begin{pmatrix} A_1 & 0 \\ * & B_1 \end{pmatrix}$ of index $q+1$. Clearly, any character of A_1 or $GL(1, q)$ multiplied by any character of B_1 or $GL(1, q)$ is a character of $GL(1, 1; q)$. If we use the linear characters of $GL(1, 1; q)$ obtainable in this way as a basis for Frobenius's method of induced

characters, we get $\frac{1}{2}(q-1)(q-2)$ irreducible characters of degree $q+1$ of GL(2, q). (See Table I.)

Finally, the linear characters of the cyclic subgroup $\begin{pmatrix} \sigma & \\ & \sigma^q \end{pmatrix}^a$ of index $q(q-1)$ induce in GL(2, q) the following representations $\Psi_{q(q-1)}{}^{(n)}$ of degree q^2-q, all of which are reducible:

$$A_1\colon (q^2-q)\,\epsilon^{na(q+1)},\quad A_2\colon 0,\quad A_3\colon 0,\quad B_1\colon \epsilon^{na}+\epsilon^{naq},$$

where $\epsilon^{q^2-1}=1$ and $n=1,2,\ldots,q-1$. But, if we form $\chi_q{}^{(o)}\chi_{q+1}{}^{(o,n)}-\chi_{q+1}{}^{(o,n)}-\psi_{q(q-1)}{}^{(n)}$, we get an irreducible character provided $n \neq$ mult. $(q+1)$. We thus have $\frac{1}{2}q(q-1)$ irreducible characters of degree $q-1$ and this completes the list since we now have in all $(q-1)(q+1)=k$ characters. They are shown in Table II.

TABLE II

Characters of GL(2, q)

Element	$\chi_1{}^{(n)}$	$\chi_q{}^{(n)}$	$\chi_{q+1}{}^{(m,n)}$	$\chi_{q-1}{}^{(n)}$
	$n=1,2,\ldots,q-1$; $\epsilon^{q-1}=1$	$n=1,2,\ldots,q-1$; $\epsilon^{q-1}=1$	$m,n=1,2,\ldots,q-1$; $m\neq n$; $(m,n)\equiv(n,m)$; $\epsilon^{q-1}=1$	$n=1,2,\ldots,q^2-2$; $n\neq$ mult. $(q+1)$; $\epsilon^{q^2-1}=1$
A_1	ϵ^{2na}	$q\epsilon^{2na}$	$(q+1)\epsilon^{(m+n)a}$	$(q-1)\epsilon^{na(q+1)}$
A_2	ϵ^{2na}	0	$\epsilon^{(m+n)a}$	$-\epsilon^{na(q+1)}$
A_3	$\epsilon^{n(a+b)}$	$\epsilon^{n(a+b)}$	$\epsilon^{ma+nb}+\epsilon^{na+mb}$	0
B_1	ϵ^{na}	$-\epsilon^{na}$	0	$-(\epsilon^{na}+\epsilon^{naq})$

The theorems of Frobenius [4], [5] mentioned in the introduction immediately give us the characters of PGL(2, q). For q even they are as in Table III. For q odd, there are in addition the two characters

$$A_1\colon 1,\ A_2\colon 1,\ A_3\colon (-1)^{a+b},\ B_1\colon (-1)^a,$$

and

$$A_1\colon q,\ A_2\colon 0,\ A_3\colon (-1)^{a+b},\ B_1\colon (-1)^{a+1}.$$

TABLE III

Characters of PGL(2, q)

Element	χ_1	χ_q	$\chi_{q+1}{}^{(n)}$	$\chi_{q-1}{}^{(n)}$
			$n=1,2,\ldots,[\frac{1}{2}(q-1)]$; $\epsilon^{q-1}=1$	$n=1,2,\ldots,[\frac{1}{2}(q+1)]$; $\epsilon^{q+1}=1$
A_1	1	q	$q+1$	$q-1$
A_2	1	0	1	-1
A_3	1	1	$\epsilon^{n(b-a)}+\epsilon^{-n(b-a)}$	0
B_1	1	-1	0	$-(\epsilon^{na}+\epsilon^{naq})$

3. The characters of GL(3, q) and PGL(3, q). The group GL(3, q) is of order $q^3(q-1)^3(q+1)(q^2+q+1)$ and each of its elements similar to one of the following types [2]:

$$A_1\colon \begin{pmatrix} \rho^a & & \\ & \rho^a & \\ & & \rho^a \end{pmatrix},\ A_2\colon \begin{pmatrix} \rho^a & & \\ 1 & \rho^a & \\ & & \rho^a \end{pmatrix},\ A_3\colon \begin{pmatrix} \rho^a & & \\ 1 & \rho^a & \\ & 1 & \rho^a \end{pmatrix},\ A_4\colon \begin{pmatrix} \rho^a & & \\ & \rho^a & \\ & & \rho^b \end{pmatrix},$$

$$A_5\colon \begin{pmatrix} \rho^a & & \\ 1 & \rho^a & \\ & & \rho^b \end{pmatrix},\ A_6\colon \begin{pmatrix} \rho^a & & \\ & \rho^b & \\ & & \rho^c \end{pmatrix},\ B_1\colon \begin{pmatrix} \rho^a & & \\ & \sigma^b & \\ & & \sigma^{bq} \end{pmatrix},\ C_1\colon \begin{pmatrix} \tau^a & & \\ & \tau^{aq} & \\ & & \tau^{aq^2} \end{pmatrix},$$

where $a \neq$ mult. (q^2+q+1) in C. The number of elements in each class and the number of classes of each type are given in Table IV. The total number of classes is $q(q-1)(q+1) = k$.

TABLE IV

Element	Number of Classes	Elements in each Class
A_1	$q-1$	1
A_2	$q-1$	$(q-1)(q+1)(q^2+q+1)$
A_3	$q-1$	$q(q-1)^2(q+1)(q^2+q+1)$
A_4	$(q-1)(q-2)$	$q^2(q^2+q+1)$
A_5	$(q-1)(q-2)$	$q^2(q-1)(q+1)(q^2+q+1)$
A_6	$\frac{1}{6}(q-1)(q-2)(q-3)$	$q^3(q+1)(q^2+q+1)$
B_1	$\frac{1}{2}q(q-1)$	$q^3(q-1)(q^2+q+1)$
C_1	$\frac{1}{3}q(q-1)(q+1)$	$q^3(q-1)^2(q+1)$

Here, as before, the permutation of the points of the underlying geometry gives us a double-transitive permutation group, in this case of degree q^2+q+1. We thus get the unit representation and an irreducible representation of degree q^2+q. The geometric entities each of which consists of a point and a line through it are also permuted by the elements of GL(3, q), and this furnishes us with a representation of degree $(q+1)(q^2+q+1)$. The orthogonality properties of group characters tell us that the character of this representation contains the unit character χ_1 once and χ_{q^2+q} twice and an irreducible character [9] of degree q^3. Multiplying each of the characters of degrees $1, q^2+q, q^3$ by each of the $q-1$ linear characters given by the powers of the determinants, we obtain $q-1$ irreducible characters of each of these degrees, as in Table V.

TABLE V

Element	$\chi_1^{(n)}$	$\chi_{q^2+q}^{(n)}$	$\chi_{q^3}^{(n)}$
A_1	ϵ^{3na}	$(q^2+q)\epsilon^{3na}$	$q^3\epsilon^{3na}$
A_2	ϵ^{3na}	$q\epsilon^{3na}$	0
A_3	ϵ^{3na}	0	0
A_4	$\epsilon^{n(2a+b)}$	$(q+1)\epsilon^{n(2a+b)}$	$q\epsilon^{n(2a+b)}$
A_5	$\epsilon^{n(2a+b)}$	$\epsilon^{n(2a+b)}$	0
A_6	$\epsilon^{n(a+b+c)}$	$2\epsilon^{n(a+b+c)}$	$\epsilon^{n(a+b+c)}$
B_1	$\epsilon^{n(a+b)}$	0	$-\epsilon^{n(a+b)}$
C_1	ϵ^{na}	$-\epsilon^{na}$	ϵ^{na}

(where $n = 1, 2, \ldots q-1$ and $\epsilon^{q-1} = 1$).

We next consider the subgroup of index q^2+q+1:

$$\mathrm{GL}(1, 2; q) = \begin{pmatrix} A_1 & 0 & 0 \\ * & A_2 & \\ * & & \end{pmatrix}.$$

It is clear that any character of A_1 (or GL(1, q)) multiplied by any character of A_2(or GL(2, q)) is a character of GL(1, 2; q). By multiplying linear characters of GL(1, q) by the characters of degree 1, q, $q+1$, $q-1$ of GL(2, q) determined in §2, we get characters of these degrees of GL(1, 2; q). These characters induce in GL(3, q) a set of characters from which we can extract $(q-1)(q-2)$ irreducible characters of degree q^2+q+1, $(q-1)(q-2)$ of degree $q(q^2+q+1)$, $\frac{1}{6}(q-1)(q-2)(q-3)$ of degree $(q+1)(q^2+q+1)$, $\frac{1}{2}q(q-1)^2$ of degree $(q-1)(q^2+q+1)$. See Table VI and Table VII.

TABLE VI

Element	$\chi_{q^2+q+1}^{(m,n)}$	$\chi_{q(q^2+q+1)}^{(m,n)}$
A_1	$(q^2+q+1)\,\epsilon^{(m+2n)a}$	$q(q^2+q+1)\,\epsilon^{(m+2n)a}$
A_2	$(q+1)\epsilon^{(m+2n)a}$	$q\epsilon^{(m+2n)a}$
A_3	$\epsilon^{(m+2n)a}$	0
A_4	$(q+1)\epsilon^{(m+n)a+nb} + \epsilon^{2na+mb}$	$(q+1)\epsilon^{(m+n)a+nb} + q\epsilon^{2na+mb}$
A_5	$\epsilon^{(m+n)a+nb} + \epsilon^{2na+mb}$	$\epsilon^{(m+n)a+nb}$
A_6	$\Sigma_{(a,b,c)}\epsilon^{ma+n(b+c)}$	$\Sigma_{(a,b,c)}\epsilon^{ma+n(b+c)}$
B_1	ϵ^{ma+nb}	$-\epsilon^{ma+nb}$
C_1	0	0

(where $m, n = 1, 2, \ldots q-1$; $m \neq n$ and $\epsilon^{q-1} = 1$).

TABLE VII

Element	$\chi_{(q+1)(q^2+q+1)}^{(l,m,n)}$	$\chi_{(q-1)(q^2+q+1)}^{(m,n)}$
	$l, m, n, = 1, 2, \ldots, q-1; l \neq m \neq n \neq l$; $\epsilon^{q-1} = 1$	$m = 1, 2, \ldots, q-1; n = 1, 2, \ldots, q^2-2$; $n \neq$ mult. $(q+1)$ $\epsilon^{q^2-1} = 1$
A_1	$(q+1)(q^2+q+1)\epsilon^{(l+m+n)a}$	$(q-1)(q^2+q+1)\epsilon^{(m+n)a(q+1)}$
A_2	$(2q+1)\epsilon^{(l+m+n)a}$	$-\epsilon^{(m+n)a(q+1)}$
A_3	$\epsilon^{(l+m+n)a}$	$-\epsilon^{(m+n)a(q+1)}$
A_4	$(q+1)\Sigma_{(l,m,n)}\epsilon^{(l+m)a+nb}$	$(q-1)\,\epsilon^{(na+mb)(q+1)}$
A_5	$\Sigma_{(l,m,n)}\epsilon^{(l+m)a+nb}$	$-\epsilon^{(na+mb)(q+1)}$
A_6	$\Sigma_{(l,m,n)}\epsilon^{la+mb+nc}$	0
B_1	0	$-\epsilon^{ma(q+1)}(\epsilon^{nb}+\epsilon^{nbq})$
C_1	0	0

By $\Sigma_{(l,m,n)}\epsilon^{(l+m)a+nb}$, we mean the symmetric function in l, m, and n which has $\epsilon^{(l+m)a+nb}$ as its typical term.

Finally, we turn to the cyclic subgroup of order $(q-1)(q^2+q+1)$:

$$\begin{pmatrix} \tau & & \\ & \tau^q & \\ & & \tau^{q^2} \end{pmatrix}^a.$$

The linear characters of this subgroup induce the following in the group GL$(3,q)$:

A_1: $q^3(q-1)^2(q+1)\,\epsilon^{na(q^2+q+1)}$, A_2: 0, A_3: 0, A_4: 0,
A_5: 0, A_6: 0, B_1: 0, C_1: $\epsilon^{na}+\epsilon^{naq}+\epsilon^{naq^2}$.

If from this character we subtract $[\chi_{q^3}^{(0)} - \chi_{q^2+q}^{(0)} + \chi_1^{(0)}]\,\chi_{(q-1)(q^2+q+1)}^{(0,n)}$, we get:

A_1: $(q-1)^2(q+1)\epsilon^{na(q^2+q+1)}$, A_2: $-(q-1)\epsilon^{na(q^2+q+1)}$, A_3: $\epsilon^{na(q^2+q+1)}$
A_4: 0, A_5: 0, A_6: 0, B_1: 0, C_1: $\epsilon^{na}+\epsilon^{naq}+\epsilon^{naq^2}$.

This is an irreducible character if $n \neq$ mult. (q^2+q+1). Since $(n) \equiv (nq) \equiv (nq^2)$, we thus get $\frac{1}{3}\,q(q-1)(q+1)$ irreducible characters of degree $(q-1)^2(q+1)$.

This completes the list of characters since we have now obtained $q(q-1)(q+1) = k$ irreducible characters.

In obtaining the characters of PGL$(3, q)$, again two cases must be distinguished: $q = 3t+1$ or $q \neq 3t+1$. The revision of classes and characters in each case is straightforward and we shall content ourselves with a list of the number of characters of each degree. (See Table VIII.)

TABLE VIII
Characters of PGL$(3, q)$

Degree	1	q^2+q	q^3	q^2+q+1	$q(q^2+q+1)$	$(q+1)\times(q^2+q+1)$	$(q-1)\times(q^2+q+1)$	$(q-1)^2\times(q+1)$
Frequency								
$q=3t+1$	3	3	3	$q-4$	$q-4$	$\frac{1}{6}(q^2-5q+10)$	$\frac{1}{2}q(q-1)$	$\frac{1}{3}(q-1)(q+2)$
$q \neq 3t+1$	1	1	1	$q-2$	$q-2$	$\frac{1}{6}(q-2)(q-3)$	$\frac{1}{2}q(q-1)$	$\frac{1}{3}q(q+1)$

4. The characters of GL(4, q) and PGL(4, q). The group GL(4, q) is of order $q^6(q-1)^4(q+1)^2(q^2+1)(q^2+q+1)$ and each of its elements is similar to one of the following twenty-two types [2]:

$$A_1:\begin{pmatrix}\rho^a&&&\\&\rho^a&&\\&&\rho^a&\\&&&\rho^a\end{pmatrix},\quad A_2:\begin{pmatrix}\rho^a&&&\\1&\rho^a&&\\&&\rho^a&\\&&&\rho^a\end{pmatrix},\quad A_3:\begin{pmatrix}\rho^a&&&\\1&\rho^a&&\\&&\rho^a&\\&&1&\rho^a\end{pmatrix},$$

$$A_4:\begin{pmatrix}\rho^a&&&\\1&\rho^a&&\\&1&\rho^a&\\&&&\rho^a\end{pmatrix},\quad A_5:\begin{pmatrix}\rho^a&&&\\1&\rho^a&&\\&1&\rho^a&\\&&1&\rho^a\end{pmatrix},\quad A_6:\begin{pmatrix}\rho^a&&&\\&\rho^a&&\\&&\rho^a&\\&&&\rho^b\end{pmatrix},$$

$$A_7:\begin{pmatrix}\rho^a&&&\\1&\rho^a&&\\&&\rho^a&\\&&&\rho^b\end{pmatrix},\quad A_8:\begin{pmatrix}\rho^a&&&\\1&\rho^a&&\\&1&\rho^a&\\&&&\rho^b\end{pmatrix},\quad A_9:\begin{pmatrix}\rho^a&&&\\&\rho^a&&\\&&\rho^b&\\&&&\rho^b\end{pmatrix},$$

$$A_{10}:\begin{pmatrix}\rho^a&&&\\1&\rho^a&&\\&&\rho^b&\\&&&\rho^b\end{pmatrix},\quad A_{11}:\begin{pmatrix}\rho^a&&&\\1&\rho^a&&\\&&\rho^b&\\&&1&\rho^b\end{pmatrix},\quad A_{12}:\begin{pmatrix}\rho^a&&&\\&\rho^a&&\\&&\rho^b&\\&&&\rho^c\end{pmatrix},$$

$$A_{13}:\begin{pmatrix}\rho^a&&&\\1&\rho^a&&\\&&\rho^b&\\&&&\rho^c\end{pmatrix},\quad A_{14}:\begin{pmatrix}\rho^a&&&\\&\rho^b&&\\&&\rho^c&\\&&&\rho^d\end{pmatrix},\quad B_1:\begin{pmatrix}\rho^a&&&\\&\rho^a&&\\&&\sigma^b&\\&&&\sigma^{bq}\end{pmatrix},$$

$$B_2:\begin{pmatrix}\rho^a&&&\\1&\rho^a&&\\&&\sigma^b&\\&&&\sigma^{bq}\end{pmatrix},\quad B_3:\begin{pmatrix}\rho^a&&&\\&\rho^b&&\\&&\sigma^c&\\&&&\sigma^{cq}\end{pmatrix},\quad C_1:\begin{pmatrix}\sigma^a&&&\\&\sigma^{aq}&&\\&&\sigma^a&\\&&&\sigma^{aq}\end{pmatrix},$$

$$C_2:\begin{pmatrix}\sigma^a&&&\\&\sigma^{aq}&&\\1&&\sigma^a&\\&1&&\sigma^{aq}\end{pmatrix},\ C_3:\begin{pmatrix}\sigma^a&&&\\&\sigma^{aq}&&\\&&\sigma^b&\\&&&\sigma^{bq}\end{pmatrix},\ D_1:\begin{pmatrix}\rho^a&&&\\&\tau^b&&\\&&\tau^{bq}&\\&&&\tau^{bq^2}\end{pmatrix},\ E_1:\begin{pmatrix}\omega^a&&&\\&\omega^{aq}&&\\&&\omega^{aq^2}&\\&&&\omega^{aq^3}\end{pmatrix}.$$

Now, we shall make use of the underlying geometry to obtain five irreducible characters. To do this, we consider the following five geometric entities: the PG(3, q); a point; a line; a point and a line through it; a point, a line through it, and a plane through the line. It will be noted that these five entities correspond to the five partitions of 4: (4), (13), (2^2), (1^22), (1^4), respectively. In fact, GL(4, q), GL(1, 3; q), GL(2, 2; q), GL(1, 1, 2; q) and GL(1, 1, 1, 1; q) are the subgroups of GL(4, q) which leave fixed one of each of these entities, respectively. Each of these sets of entities will be permuted by the elements of

TABLE IX

Characters of GL(4 q) and PGL(4, q)

Element	Unit	Point	$q(q^2+q+1)$	Line	$q^2(q^2+1)$	Point–line	$q^3(q^2+q+1)$	Point–Line–Plane	q^6
A_1	1	$(q+1)(q^2+1)$	$q(q^2+q+1)$	$(q^2+1)(q^2+q+1)$	$q^2(q^2+1)$	$(q+1)(q^2+1)(q^2+q+1)$	$q^3(q^2+q+1)$	$(q+1)^2(q^2+1)(q^2+q+1)$	q^6
A_2	1	q^2+q+1	q^2+q	$2q^2+q+1$	q^2	q^3+3q^2+2q+1	q^3	$3q^3+5q^2+3q+1$	0
A_3	1	$q+1$	q	q^2+q+1	q^2	q^2+2q+1	0	$2q^2+3q+1$	0
A_4	1	$q+1$	q	$q+1$	0	$2q+1$	0	$3q+1$	0
A_5	1	1	0	1	0	1	0	1	0
A_6	1	q^2+q+2	q^2+q+1	$2q^2+2q+2$	q^2+q	q^3+4q^2+4q+3	q^3+q^2+q	$4q^3+8q^2+8q+4$	q^2
A_7	1	$q+2$	$q+1$	$2q+2$	q	$4q+3$	q	$8q+4$	0
A_8	1	2	1	2	0	3	0	4	0
A_9	1	$2q+2$	$2q+1$	q^2+2q+3	q^2+1	$2q^2+6q+4$	q^2+2q	$6q^2+12q+6$	q^2
A_{10}	1	$q+2$	$q+1$	$q+3$	1	$3q+4$	q	$6q+6$	0
A_{11}	1	2	1	3	1	4	0	6	0
A_{12}	1	$q+3$	$q+2$	$2q+4$	$q+1$	$5q+7$	$2q+1$	$12q+12$	q
A_{13}	1	3	2	4	1	7	1	12	0
A_{14}	1	4	3	6	2	12	3	24	1
B_1	1	$q+1$	q	2	$-q+1$	$q+1$	-1	0	$-q$
B_2	1	1	0	2	1	1	-1	0	0
B_3	1	2	1	2	0	2	-1	0	-1
C_1	1	0	-1	q^2+1	q^2+1	0	$-q^2$	0	q^2
C_2	1	0	-1	1	1	0	0	0	0
C_3	1	0	-1	2	2	0	-1	0	1
D_1	1	1	0	0	-1	0	0	0	1
D_2	1	0	-1	0	0	0	1	0	-1

GL(4, q) and in this way permutation representations of degree 1, $(q+1)\cdot(q^2+1)$, $(q^2+1)(q^2+q+1)$, $(q+1)(q^2+1)(q^2+q+1)$ and $(q+1)^2(q^2+1)(q^2+q+1)$ will be obtained. All except the first of these five characters are reducible, but they can be combined to give five irreducible characters as follows [9]:

$$1 = 1;\ (q+1)(q^2+1) - 1 = q(q^2+q+1);$$
$$(q^2+1)(q^2+q+1) - (q+1)(q^2+1) = q^2(q^2+1);$$
$$(q+1)(q^2+1)(q^2+q+1) - (q^2+1)(q^2+q+1) - (q+1)(q^2+1) + 1 = q^3(q^2+q+1);$$
$$(q+1)^2(q^2+1)(q^2+q+1) - 3(q+1)(q^2+1)(q^2+q+1) + (q^2+1)(q^2+q+1) + 2(q+1)(q^2+1) - 1 = q^6.$$

Multiplication of each of these characters by the $q-1$ linear characters given by the powers of the determinants gives $q-1$ irreducible characters of each of these degrees. Table IX lists the basic characters and shows the "fixed entity" situation.

We next consider characters induced by those of subgroup GL(1, 3; q) of index $(q+1)(q^2+1)$. In a manner analogous to those obtained of GL(3, q) from GL(1, 2; q), we get irreducible characters of the degrees and frequencies[3] shown in Table X:

TABLE X

Degree	Frequency
$(q+1)(q^2+1)$	$(q-1)(q-2)$
$q(q+1)^2(q^2+1)$	$(q-1)(q-2)$
$q^2(q+1)(q^2+1)$	$(q-1)(q-2)$
$(q+1)(q^2+1)(q^2+q+1)$	$\frac{1}{2}(q-1)(q-2)(q-3)$
$q(q+1)(q^2+1)(q^2+q+1)$	$\frac{1}{2}(q-1)(q-2)(q-3)$
$(q+1)^3(q^2+1)(q^2+q+1)$	$\frac{1}{24}(q-1)(q-2)(q-3)(q-4)$
$(q-1)(q+1)(q^2+1)(q^2+q+1)$	$\frac{1}{4}q(q-1)^2(q-2)$
$(q-1)^2(q+1)^2(q^2+1)$	$\frac{1}{8}q(q-1)^2(q+1)$

In the same way, the subgroup GL(2, 2; q) yields the irreducible characters shown in Table XI:

TABLE XI

Degree	Frequency
$(q^2+1)(q^2+q+1)$	$\frac{1}{2}(q-1)(q-2)$
$q^2(q^2+1)(q^2+q+1)$	$\frac{1}{2}(q-1)(q-2)$
$q(q^2+1)(q^2+q+1)$	$(q-1)(q-2)$
$(q-1)(q^2+1)(q^2+q+1)$	$\frac{1}{2}q(q-1)^2$
$q(q-1)(q^2+1)(q^2+q+1)$	$\frac{1}{2}q(q-1)^2$
$(q-1)^2(q^2+1)(q^2+q+1)$	$\frac{1}{4}q(q-1)(q+1)(q-2)$.

[3]The actual characters of GL(4, q) with a more detailed account of the methods are available in [10].

As a bi-product of the set of characters of degree $(q-1)^2(q^2+1)(q^2+q+1)$ we obtain $\frac{1}{2}q(q-1)$ characters of this degree each of which is the sum of two irreducible characters which are not among those that we have already obtained. Let us denote them by $\chi^{(n)}$, $n = 1, 2, \ldots, \frac{1}{2}q(q-1)$.

Finally, the linear characters of the cyclic subgroup of order q^4-1,

$$\begin{pmatrix} \omega & & & \\ & \omega^{q} & & \\ & & \omega^{q^2} & \\ & & & \omega^{q^3} \end{pmatrix}^{a},$$

induce in $\mathrm{GL}(4, q)$ a set of characters of degree $q^6(q-1)^3(q+1)(q^2+q+1)$. Each of these is reducible, but by a suitable use of the characters already obtained, i.e., by multiplication, addition and subtraction, a set of $\frac{1}{4}q^2(q-1)(q+1)$ irreducible characters of degree $(q-1)^3(q+1)(q^2+q+1)$ can be extracted from them. Again there is a bi-product: $\frac{1}{2}q(q-1)$ pseudocharacters of degree $(q-1)^3(q+1)(q^2+q+1)$ each of which is the difference of two irreducible characters. Denote them by $\psi^{(n)}$. Then, if the proper correlation is made be-

TABLE XII

Characters of PGL(4, q)

Degrees	Frequencies		
	$q = 4t$ or $4t+2$	$q = 4t+1$	$q = 4t+3$
1	1	4	2
(10) (111)	1	4	2
$(10)^2(101)$	1	4	2
$(10)^3(111)$	1	4	2
$(10)^6$	1	4	2
(11)(101)	$1-2$	$1-5$	$1-3$
$(10)\ (11)^2(101)$	$1-2$	$1-5$	$1-3$
$(10)^3(11)\ (101)$	$1-2$	$1-5$	$1-3$
(11) (101) (111)	$\frac{1}{2}(1-2)(1-3)$	$\frac{1}{2}(1-6-13)$	$\frac{1}{2}(1-3)^2$
(10)(11)(101)(111)	$\frac{1}{2}(1-2)(1-3)$	$\frac{1}{2}(1-6-13)$	$\frac{1}{2}(1-3)^2$
$(11)^2(101)(111)$	$\frac{1}{24}(1-2)(1-3)(1-4)$	$\frac{1}{24}(1-5)(1-49)$	$\frac{1}{24}(1-3)(1-6-11)$
$(1-1)(11)(101)(111)$	$\frac{1}{4}(10)(1-1)(1-2)$	$\frac{1}{4}(1-1)^2$	$\frac{1}{4}(1-1)^2$
$(1-1)^2(11)^2(101)$	$\frac{1}{2}(10)(1-1)(11)$	$\frac{1}{2}(10)(1-1)(11)$	$\frac{1}{2}(10)(1-1)(11)$
(101)(111)	$\frac{1}{2}(1-2)$	$1-3$	$1-2$
$(10)^3(101)(111)$	$\frac{1}{2}(1-2)$	$1-3$	$1-2$
(10) (101)(111)	$1-2$	$2-6$	$2-4$
$(1-1)(101)(111)$	$\frac{1}{2}(10)(1-1)$	$\frac{1}{2}(1-1)^2$	$\frac{1}{2}(1-1)^2$
$(10)(1-1)(101)(111)$	$\frac{1}{2}(10)(1-1)$	$\frac{1}{2}(1-1)^2$	$\frac{1}{2}(1-1)^2$
$(1-1)^2(101)(111)$	$\frac{1}{2}(10)(11)(1-2)$	$\frac{1}{2}(1-1)(10-3)$	$\frac{1}{2}(11)(1-2-1)$
$(1-1)^2(111)$	$\frac{1}{2}(10)$	$1-1$	10
$(10)^2(1-1)^2(111)$	$\frac{1}{2}(10)$	$1-1$	10
$(1-1)^3(11)(111)$	$\frac{1}{4}(10)^2(11)$	$\frac{1}{4}(1-1)(11)^2$	$\frac{1}{4}(1-1)(11)^2$

tween the $\chi^{(n)}$'s and the $\psi^{(n)}$'s, it turns out that $\frac{1}{2}(\chi^{(n)}+\psi^{(n)})$ and $\frac{1}{2}(\chi^{(n)}-\psi^{(n)})$ are irreducible characters. In this way we obtain $\frac{1}{2}\, q(q-1)$ irreducible characters of each of the degrees $q^2(q-1)^2(q^2+q+1)$ and $(q-1)^2(q^2+q+1)$. This completes the character list since we have now obtained $q^4-q=k$ of them.

In cutting down the characters of GL(4, q) to get those of PGL(4, q), three cases are distinct: q even, $q=4t+1$, $q=4t+3$. Table XII gives the degrees and frequencies in each of these cases. For convenience in notation, we shall mean by $\frac{1}{2}$ (10-11), for example, $\frac{1}{2}$ (q^3-q+1), etc.

References

[1] H. W. Brinkmann, Bull. Amer. Math. Soc., vol. 27 (1921), 152.

[2] L. E. Dickson, *Linear Groups in Galois Fields* (Leipzig, 1901).

[3] G. Frobenius, *Über Gruppencharaktere*, Berliner Sitz. (1896), 985.

[4] ———, *Über die Darstellung der endlichen Gruppen durch Lineare Substitutionen*, Berliner Sitz. (1897), 994.

[5] ———, *Über Relationen zwischen den Charakteren einer Gruppe und denen ihrer Untergruppen*, Berliner Sitz. (1898), 501.

[6] H. Jordan, *Group-Characters of Various Types of Linear Groups*, Amer. J. of Math., vol. 29 (1907), 387.

[7] I. Schur, *Untersuchungen über die Darstellung der endlichen Gruppen durch Gebrochene Lineare Substitutionen*, J. für Math., vol. 132 (1907), 85.

[8] A. Speiser, *Die Theorie der Gruppen von endlicher Ordnung* (Berlin, 1937).

[9] R. Steinberg, *A Geometric Approach to the Representations of the Full Linear Group over a Galois Field*, submitted to Trans. Amer. Math. Soc.

[10] ———, *Representations on the Linear Fractional Groups*, Thesis, University of Toronto Library.

University of California at Los Angeles

DIRECT THEOREMS ON METHODS OF SUMMABILITY II

G. G. LORENTZ

1. Introduction

1.1. This paper is a continuation of the papers of the author [14], [15]. We begin by recapitulating the main definitions. If $\{n_\nu\}$ is an increasing sequence of positive integers, the value of the *characteristic or the counting function* $\omega(n)$ of $\{n_\nu\}$ is, for any $n \geqslant 0$, the number of n_ν satisfying the inequality $n_\nu \leqslant n$. Suppose that A is a linear method of summation corresponding to the transformation

$$\sigma_m = \sum_{n=0}^{\infty} a_{mn} s_n \qquad (m = 0, 1, \ldots). \tag{1.1(1)}$$

In what follows, $\Omega(n)$ is always a non-decreasing positive function defined for all real $n \geqslant 0$ and tending to $+\infty$ with n. A function $\Omega(n)$ is a *summability function of the first kind of a method* A if all real bounded sequences s_n such that $s_n = 0$ except for a sequence $\{n_\nu\}$ of values of n whose counting function $\omega(n) \leqslant \Omega(n)$, $n \geqslant 0$, are A-summable. $\Omega(n)$ is a *summability function of the second kind of a method* A if $S_n = s_0 + s_1 + \ldots + s_n = O(\Omega(n))$ implies that s_n is A-summable.

In [15] we have given necessary and sufficient conditions for summability functions of an arbitrary method A and have found all summability functions of some special methods. Here in §2 and §3 we solve the last problem for the Riesz and Abel methods $R(\lambda_n,\kappa)$, $\kappa > 0$ and $A(\lambda_n)$ (for the properties of these methods compare Hardy and Riesz [6], Hardy [5]). We have had to make some hypotheses on the regularity of the sequence λ_n (which are in most cases very modest). In §4 we discuss summability functions for absolute summability. Theorem 6 gives necessary and sufficient conditions for absolute summability functions, Theorem 7 describes methods which possess such functions. We also determine all absolute summability functions for some special methods. Thus for the Cesàro methods C_α, $\alpha > 0$ they are given by the condition $\sum n^{-1-\beta}\Omega(n) < +\infty$ ($\beta = \alpha$ for $\alpha \leqslant 1$, $\beta = 1$ for $\alpha \geqslant 1$) in contrast to the condition $\Omega(n) = o(n)$ which describes ordinary summability functions of C_α. Finally, in §5 we give applications of theorems of this and the previous papers. Of these we note Theorem 10, whose application is a good way to show that certain Tauberian conditions are the best possible of their kind.

Received February 13, 1950.

2. Summability functions of Riesz and Abel methods. Case when $\Delta\lambda_n$ is increasing

2.1. Let $0 = \lambda_0 < \lambda_1 < \lambda_2 < \dots, \lambda_n \to \infty$ be a given sequence and $\kappa > 0$. *A* series $\sum u_n$, or the sequence s_n of its partial sums, is $R(\lambda_n, \kappa)$ summable to s if

$$2.1(1) \qquad v^{-\kappa} \sum_{\lambda_n \leqslant v} (v - \lambda_n)^\kappa u_n$$

converges to s for $v \to \infty$. And $\sum u_n$ is $A(\lambda_n)$ -summable to s, if

$$2.1(2) \qquad \sigma(x) = \sum_{n=0}^{\infty} e^{-\lambda_n x} u_n \to s, \qquad x \to 0+.$$

We shall find it convenient to extend the definition of λ_n also to non-integral values of n and to consider a monotone continuous function $\lambda(\omega)$, $\omega \geqslant 0$ such that $\lambda(n) = \lambda_n$. Then we can write 2.1(1) in the form

$$2.1(3) \qquad \begin{aligned} \sigma(\omega) &= \lambda(\omega)^{-\kappa} \sum_{n \leqslant \omega} (\lambda(\omega) - \lambda_n)^\kappa u_n \\ &= \lambda(\omega)^{-\kappa} \sum_{n \leqslant n_0 - 1} \left\{(\lambda(\omega) - \lambda_n)^\kappa - (\lambda(\omega) - \lambda_{n+1})^\kappa\right\} s_n + \lambda(\omega)^{-\kappa}(\lambda(\omega) - \lambda_{n_0})^\kappa s_{n_0}, \end{aligned}$$

where $n_0 = [\omega]$. On the other hand, the expression 2.1(2) is equivalent to

$$2.1(4) \qquad \sigma(x) = \sum_{n=0}^{\infty} (e^{-\lambda_n x} - e^{-\lambda_{n+1} x}) s_n$$

for any $A(\lambda_n)$ -summable sequence s_n (see for instance [13, Theorem 10]).

In the sequel we seek to find all summability functions of the methods $R(\lambda_n, \kappa)$, $A(\lambda_n)$ in a simpler form than that given by general theorems [15, §2]. We first make the following remark. *Any of the methods* $R(\lambda_n, \kappa)$, $\kappa > 0$, $A(\lambda_n)$ *possesses summability functions if and only if*

$$2.1(5) \qquad \Delta\lambda_n/\lambda_n \to 0 \ \text{ or } \ \lambda_{n+1}/\lambda_n \to 1 \qquad (\Delta\lambda_n = \lambda_{n+1} - \lambda_n).$$

In fact, if the method $R(\lambda_n, \kappa)$ has summability functions, the coefficients of the transformation 2.1(3) must converge uniformly to zero for $\omega \to \infty$ by [14, Theorem 8*]. In particular the last coefficient converges to zero, and this gives 2.1(5). And if 2.1(5) is true, the coefficients in 2.1(4) converge uniformly to 0:

$$e^{-\lambda_n x}(1 - e^{-\Delta\lambda_n x}) \leqslant C_1 e^{-\lambda_n x} \Delta\lambda_n x \leqslant C_2 \Delta\lambda_n/\lambda_n \to 0,$$

since $e^{-u}u$ is bounded for $u \geqslant 0$. Since $R(\lambda_n, \kappa) \subset A(\lambda_n)$ for $\kappa > 0$ [6, p. 39], the proof is complete.

2.2. To obtain further results we suppose some regularity of the sequence λ_n. In this section we shall suppose that $\Delta\lambda_n$ is increasing. A first consequence of this hypothesis together with 2.1(5) is that $\lambda_n/\Delta\lambda_n = O(n)$. For

$$\Delta\left(\frac{\lambda_n}{\Delta\lambda_n} - n\right) = \left(\frac{\Delta\lambda_n}{\Delta\lambda_{n+1}} - 1\right) - \frac{\lambda_n \Delta^2\lambda_n}{\Delta\lambda_n \Delta\lambda_{n+1}} \leqslant 0,$$

and thus $\lambda_n/\Delta\lambda_n - n$ is decreasing. Therefore, $\lambda_n/\Delta\lambda_n \leqslant n + C$ for some constant C. Theorems 1 and 2 below give full information about the summability functions of the first and the second kind. In Theorem 1 we suppose that $\Delta\lambda_n/\lambda_n \to 0$ (which is no restriction because of 2.1(5)), in Theorem 2 slightly more, namely that $\Delta\lambda_n/\lambda_n$ decreases to 0.

THEOREM 1. *If $\Delta\lambda_n/\lambda_n$ converges to zero and $\Delta\lambda_n$ increases, all summability functions (of the first kind) of the methods* $\mathrm{R}(\lambda_n, \kappa)$, $\kappa > 0$ *and* $\mathrm{A}(\lambda_n)$, *and only these functions, are given by*

2.2(1) $$\Omega(n) = o(\lambda_n/\Delta\lambda_n).$$

Proof. (a) *Every function $\Omega(n)$ satisfying* 2.2(1) *is a summability function of the method* $\mathrm{R}(\lambda_n, \kappa)$, $0 < \kappa \leqslant 1$. We have to show that 2.2(1) implies that $A(\omega, \Omega) \to 0$ for $\omega \to \infty$ [15, 2.3]. We recall that for a method of summation defined by $s = \lim_{\omega\to\infty} \sum_{n=1}^{\infty} a_n(\omega)s_n$ and a function $\Omega(n)$, $A(\omega, \Omega)$ is the least upper bound of $\sum_{\nu=1}^{\infty}|a_{n_\nu}(\omega)|$ for all sequences n_ν with the counting function $\leqslant \Omega(n)$. Because of 2.1(5) we may disregard the last coefficient in 2.1(3). For $n \leqslant n_0 - 1$ the coefficient

$$a_n(\omega) = -\lambda(\omega)^{-\kappa}\Delta(\lambda(\omega)-\lambda_n)^{\kappa} = \kappa\lambda(\omega)^{-\kappa}(\lambda(\omega)-\lambda'_n)^{\kappa-1}\Delta\lambda_n$$

(λ'_n is between λ_n and λ_{n+1}) is increasing with n. Therefore,

$$A(\omega, \Omega) \leqslant \sum_{n_0-\Omega(\omega)\leqslant n\leqslant n_0-1} a_n(\omega) \leqslant \lambda(\omega)^{-\kappa}[\lambda(\omega)-\lambda(n_0-\Omega(\omega))]^{\kappa}$$
$$\leqslant C\left[\frac{\Delta\lambda_{n_0+1}}{\lambda_{n_0+1}}(\Omega(\omega)+2)\right]^{\kappa} \to 0$$

by 2.1(5) and 2.2(1). This proves (a).

(b) *Any summability function of the method* $\mathrm{A}(\lambda_n)$ *satisfies* 2.2(1). Suppose that 2.2(1) does not hold, then for some $\delta > 0$ and an infinity of n, $\Omega(n) \geqslant \delta\lambda_n/\Delta\lambda_n$. For these n define the integer n_1 by

2.2(2) $$\lambda_{n_1} \leqslant (1+\delta)\lambda_n < \lambda_{n_1+1}.$$

For a fixed n of the above kind, we denote by $\Omega_1(\nu)$ the counting function of the set of integers ν defined by $n \leqslant \nu < n_1$. We have

$$n_1 - n \leqslant (\lambda_{n_1}-\lambda_n)/\Delta\lambda_n \leqslant \delta\lambda_n/\Delta\lambda_n,$$

and therefore $\Omega_1(n_1) \leqslant \Omega(n)$. Thus $\Omega_1(u) \leqslant \Omega(u)$ in $n \leqslant u < n_1$, and since Ω_1 is constant outside of this interval, the same inequality holds for all u. Therefore for the function $A(x, \Omega)$ of the method $\mathrm{A}(\lambda_n)$ we have

$$A(x, \Omega) \geqslant \sum_{n\leqslant\nu<n_1}(e^{-\lambda_\nu x}-e^{-\lambda_{\nu+1}x}) = e^{-\lambda_n x}-e^{-\lambda_{n_1}x}$$
$$= e^{-\lambda'_n x}x(\lambda_{n_1}-\lambda_n)$$

for some λ'_n between λ_n and λ_{n_1}. Here

$$\lambda_{n_1}-\lambda_n = \lambda_{n_1+1}-\lambda_n + o(\lambda_n) \geqslant \delta\lambda_n + o(1) \geqslant \tfrac{1}{2}\delta\lambda_n$$

for large n. Choosing $x_n = \lambda_n^{-1}$, we obtain $\lambda'_n x_n \leqslant 1 + \delta$ and therefore

$$A(x_n, \Omega) \geqslant \tfrac{1}{2}\delta e^{-(1+\delta)} = \text{const.} > 0,$$

so that $A(x, \Omega)$ does not tend to zero for $x \to \infty$, which proves (b) by [15, 2.3].

From (a) and (b) the theorem follows in virtue of the inclusions $\mathrm{R}(\lambda_n, \kappa) \subset \mathrm{R}(\lambda_n, \kappa') \subset \mathrm{A}(\lambda_n)$, $0 < \kappa < \kappa'$.

2.3. We now treat summability functions of the second kind.

THEOREM 2. *If $\Delta\lambda_n/\lambda_n$ decreases to 0 and $\Delta\lambda_n$ increases,* (i) *all summability functions of the second kind of the methods* $\mathrm{R}(\lambda_n, \kappa)$, $\kappa \geqslant 1$ *and* $\mathrm{A}(\lambda_n)$ *and only these are given by*

2.3(1) $$\Omega(n) = o(\lambda_n/\Delta\lambda_n).$$

(ii) *For* $\mathrm{R}(\lambda_n, \kappa)$, $0 < \kappa < 1$ *the condition is*

2.3(2) $$\Omega(n) = o(\lambda_n/\Delta\lambda_n)^{\kappa}.$$

Proof. (a) *If* 2.3(1) *holds, then* $\Omega(n)$ *is a summability function of the second kind of* $\mathrm{R}(\lambda_n, 1)$. From this (i) will follow by Theorem 1. By [15, 2.3] we have to show that if 2.3(1) holds, and $a_n(\omega)$ is the coefficient of s_n in the transformation 2.1(3) for $\kappa = 1$, then

2.3(3) $$\Delta(\omega, \Omega) = \sum_{\nu=0}^{\infty} \Omega(\nu) |\Delta a_\nu(\omega)| \to 0.$$

We have $a_\nu(\omega) = \Delta\lambda_\nu/\lambda(\omega)$ for $\nu \leqslant n_0-1$, $a_{n_0}(\omega) = (\lambda(\omega)-\lambda_{n_0})/\lambda(\omega)$ and $a_\nu(\omega) = 0$ for $\nu > n_0$. The last non-vanishing term of the sum 2.3(3) with $\nu = n_0$ converges to 0 because $\Delta\lambda_n/\lambda_n \to 0$. Therefore, 2.3(3) is equivalent to

2.3(4) $$\lambda(n)^{-1} \sum_{\nu=0}^{n} \Omega(\nu)\,\Delta^2\lambda_\nu \to 0 \qquad \text{for } n \to \infty.$$

With $\Delta\lambda_\nu/\lambda_\nu$ also $\lambda_{\nu+1}/\lambda_\nu$ is decreasing, and so $\lambda_\nu\lambda_{\nu+2} \leqslant \lambda_{\nu+1}^2$ and

2.3(5) $$\lambda_\nu\Delta^2\lambda_\nu = \lambda_\nu\lambda_{\nu+2} - 2\lambda_\nu\lambda_{\nu+1} + \lambda_\nu^2 \leqslant (\lambda_{\nu+1}-\lambda_\nu)^2 = (\Delta\lambda_\nu)^2,$$

$$\lambda_n^{-1}\sum_{\nu=0}^{n} \lambda_\nu\Delta^2\lambda_\nu/\Delta\lambda_\nu \leqslant \lambda_n^{-1}\sum_{\nu=0}^{n}\Delta\lambda_\nu = 1.$$

By a variant of the theorem of Silverman-Toeplitz we now see that

$$\lambda_n^{-1}\sum_{0}^{n} \Omega(\nu)\,\Delta^2\lambda_\nu = \lambda_n^{-1}\sum_{\nu=0}^{n} \frac{\lambda_\nu\Delta^2\lambda_\nu}{\Delta\lambda_\nu}\,\Omega(\nu)\frac{\Delta\lambda_\nu}{\lambda_\nu} \to 0,$$

if $\Omega(\nu)\,\Delta\lambda_\nu/\lambda_\nu \to 0$.

(b) *We prove* (ii). For $0 < \kappa < 1$ the necessary and sufficient condition is again 2.3(3), where $a_\nu(\omega)$ is defined by the transformation 2.1(3). Con-

sidering the last non-vanishing term we see that $\Omega(n)(\Delta\lambda_n/\lambda_n)^\kappa \to 0$, that is 2.3(2) is necessary. Let 2.3(2) be true. Then 2.3(3) is equivalent to

$$2.3(4) \qquad S = \lambda(\omega)^{-\kappa} \sum_{n=0}^{\omega_1} \Omega(n) \left| \Delta^2(\lambda(\omega)-\lambda_n)^\kappa \right| \to 0,$$

where ω_1 is some integer of the form $\omega_1 = \omega - p$, and p is constant. It will be sufficient to take $p \geqslant 5$.

We have, if $0 \leqslant c < b < a$ and $a - 2b + c \leqslant 0$,

$$\begin{aligned} a^\kappa - 2b^\kappa + c^\kappa &= c^\kappa - (2b-a)^\kappa + a^\kappa - 2b^\kappa + (2b-a)^\kappa \\ &= \kappa(a-2b+c)\xi^{\kappa-1} + \kappa(\kappa-1)\,(b-a)^2\eta^{\kappa-2}, \end{aligned}$$

where $c < \xi < 2b - a < b$, $c < \eta < a$. Applying this to S, we obtain

$$S \leqslant C_1S_1 + C_2S_2,$$

$$2.3(6) \qquad \begin{cases} S_1 = \lambda(\omega)^{-\kappa} \displaystyle\sum_{n=0}^{\omega_1} \Omega(n)\, \Delta^2\lambda_n \left|\lambda(\omega) - \lambda'_n\right|^{\kappa-1}, \\ S_2 = \lambda(\omega)^{-\kappa} \displaystyle\sum_{n=0}^{\omega_1} \Omega(n)\, (\Delta\lambda_n)^2 \left|\lambda(\omega) - \lambda''_n\right|^{\kappa-2}, \end{cases}$$

where λ'_n and λ''_n are between λ_n and λ_{n+2}. If μ_n is such that $-\Delta(\lambda(\omega)-\lambda_n)^\kappa = \kappa(\lambda(\omega)-\mu_n)^{\kappa-1}\Delta\lambda_n$, $\lambda_n < \mu_n < \lambda_{n+1}$, we have

$$\frac{\lambda(\omega)-\lambda'_n}{\lambda(\omega)-\mu_n} = 1 - \frac{\lambda'_n - \mu_n}{\lambda(\omega)-\mu_n} \geqslant 1 - \frac{\lambda_{n+2}-\lambda_n}{\lambda_{n+5}-\lambda_{n+1}} \geqslant \tfrac{1}{2},$$

and therefore, using again 2.3(5),

$$\begin{aligned} S_1 &\leqslant C_3\lambda(\omega)^{-\kappa} \sum_{n=0}^{\omega_1} \Omega(n) \frac{\Delta^2\lambda_n}{\Delta\lambda_n} [(\lambda(\omega)-\lambda_n)^\kappa - (\lambda(\omega)-\lambda_{n+1})^\kappa] \\ &\leqslant C_3\lambda(\omega)^{-\kappa} \sum_{n=0}^{\omega_1} \Omega(n) \frac{\Delta\lambda_n}{\lambda_n} [(\lambda(\omega)-\lambda_n)^\kappa - (\lambda(\omega)-\lambda_{n+1})^\kappa]. \end{aligned}$$

We may regard this as a transformation of the sequence $\Omega(n)\Delta\lambda_n/\lambda_n$ and obtain as before $S_1 \to 0$ for $\omega \to \infty$.

To deal with S_2, 2.3(1) will not be enough and we need 2.3(2) in full. We have

$$\begin{aligned} S_2 &= o(1)\lambda(\omega)^{-\kappa} \sum_{n=0}^{\omega_1} \lambda_n^\kappa(\Delta\lambda_n)^{1-\kappa}(\lambda(\omega)-\lambda_n)^{\kappa-2}\Delta\lambda_n \\ &\leqslant o(1)\,(\Delta\lambda_{n_0})^{1-\kappa} \sum_{n=0}^{\omega_1} (\lambda(\omega)-\lambda_n)^{\kappa-2}\Delta\lambda_n. \end{aligned}$$

As before, it is easy to see that $(\lambda(\omega)-\lambda_n)^{\kappa-2}\Delta\lambda_n = O(\Delta(\lambda(\omega)-\lambda_n)^{\kappa-1})$, and therefore

$$S_2 = o(1)\,(\Delta\lambda_{n_0})^{1-\kappa}[(\lambda(\omega)-\lambda(\omega_1+1))^{\kappa-1} - \lambda(\omega)^{\kappa-1}].$$

Since $\Delta\lambda_n/\lambda_n$ is decreasing, $1 \leqslant \Delta\lambda_{n+1}/\Delta\lambda_n \leqslant \lambda_{n+1}/\lambda_n \to 1$ and so $\Delta\lambda_{n+1}/\Delta\lambda_n \to 1$ for $n \to \infty$. But this implies $[\lambda(\omega)-\lambda(\omega_1+1)]/\Delta\lambda_{n_0} = O(1)$ and $S_2 = o(1)$. Therefore $S \to 0$ and the proof of the theorem is complete.

3. Riesz and Abel methods. Case when $\Delta\lambda_n$ is decreasing

3.1. If $\Delta\lambda_n$ is decreasing, the condition 2.1(5) is automatically fulfilled. By the argument used in §2.2 it is seen that we even have $\lambda_n/\Delta\lambda_n \geqslant Cn$ for some constant $C > 0$.

THEOREM 3. *If $\Delta\lambda_n$ is decreasing, all functions $\Omega(n) = o(n)$ are summability functions of the methods* $\mathrm{R}(\lambda_n, \kappa)$, $\kappa > 0$ *and* $\mathrm{A}(\lambda_n)$.

Proof. It is sufficient to consider $\mathrm{R}(\lambda_n, \kappa)$ for $0 < \kappa \leqslant 1$. We prove that $A(\omega, \Omega) \to 0$, if $\Omega(n) = o(n)$. Choose an $\epsilon > 0$ and break the matrix $A = (a_n(\omega))$ of $\mathrm{R}(\lambda_n, \mathrm{K})$ into the parts $A' = (a'_n(\omega))$, $A'' = (a''_n(\omega))$, where

$$a'_n(\omega) = \begin{cases} a_n(\omega) & \text{for } 0 \leqslant n \leqslant \omega_1 - 1, \\ 0 & \text{for } n > \omega_1 - 1; \end{cases}$$

$$a''_n(\omega) = \begin{cases} 0 & \text{for } 0 \leqslant n \leqslant \omega_1 - 1, \\ a_n(\omega) & \text{for } n > \omega_1 - 1, \end{cases}$$

and ω_1 is defined by $\lambda(\omega_1) = (1-\epsilon)\lambda(\omega)$. Clearly,

$$A(\omega, \Omega) \leqslant A'(\omega, \Omega) + A''(\omega, \Omega).$$

For $n \leqslant \omega_1 - 1$ we have, with some λ'_n between λ_n and λ_{n+1}

$$\begin{aligned} a'_n(\omega) &= \kappa\lambda(\omega)^{-\kappa}(\lambda(\omega) - \lambda'_n)^{\kappa-1}\Delta\lambda_n \\ &= \kappa\left(\frac{\lambda(\omega)}{\lambda(\omega)-\lambda'_n}\right)^{1-\kappa}\frac{\Delta\lambda_n}{\lambda(\omega)} \leqslant \frac{\kappa}{\epsilon^{1-\kappa}}\,\frac{\Delta\lambda_n}{\lambda(\omega)} = \alpha_n(\omega), \end{aligned}$$

say. We put $\alpha_n(\omega) = 0$ for $n > \omega_1 - 1$. These $\alpha_n(\omega)$ are positive, decreasing and have uniformly bounded sums $\sum_n \alpha_n(\omega)$. Therefore, $A'(\omega, \Omega) \to 0$, by [15, Theorem 7]. On the other hand,

$$\begin{aligned} A''(\omega, \Omega) &\leqslant \sum_{n=0}^{\infty} a''_n(\omega) \leqslant \lambda(\omega)^{-\kappa}(\lambda(\omega)-\lambda(\omega_1-1))^{\kappa} \\ &= (1-(1-\epsilon) + o(1))^{\kappa} = (\epsilon + o(1))^{\kappa}. \end{aligned}$$

Therefore $\overline{\lim}_{\omega\to\infty} A(\omega, \Omega) \leqslant \epsilon^{\kappa}$; and since $\epsilon > 0$ was arbitrary, $\lim A(\omega, \Omega) = 0$, q.e.d.

THEOREM 4. *If $\Delta\lambda_n$ decreases, all functions $\Omega(n) = o(n)$ are summability functions of the second kind of the methods* $\mathrm{A}(\lambda_n)$ *and* $\mathrm{R}(\lambda_n, \kappa)$, $\kappa \geqslant 1$.

Proof. It is sufficient to consider $\mathrm{R}(\lambda_n, 1)$. The assertion is then $\mathrm{R}(\lambda_n, 1) \supset \mathrm{C}_1$, and this is a theorem of Cesàro [5, p. 58].

Theorems 3 and 4 give only sufficient conditions, but it is clear that they

may not be improved, since $\Omega(n) = n$ is not a summability function for any regular method. On the other hand, summability functions which do not satisfy $\Omega(n) = o(n)$ may exist. For instance the method R $(\log n, 1)$, which is equivalent to the method of logarithmic means, possesses summability functions $\Omega(n)$ such that $\Omega(n) \neq o(\varphi(n))$ provided $\varphi(n)$ has the property $\varphi(n) = o(n \log n)$.

3.2. Now we shall show that in case $0 < \kappa < 1$ the condition for a summability function $\Omega(n)$ of the second kind is again 2.3(2). But for this result we require a much greater amount of regularity of $\lambda(n)$ than up to now. However, any function $\lambda(n)$ which is a product of powers of n and iterated logarithms satisfies our conditions.

THEOREM 5. *If for all large real n*

(α) $\lambda(n + h) - \lambda(n)$ *is decreasing for any fixed* $h > 0$,

(β) $\lambda(\log n)/\lambda(n)$ *is decreasing,*

(γ) $\Delta\lambda_n/\Delta\lambda_{2n} \leqslant M$,

then the general form of a summability function $\Omega(n)$ of the second kind of the method R(λ_n, κ), $0 < \kappa < 1$ *is* 2.3(2).

For instance, if $\lambda_n = \log\log n$, the conditions are satisfied and we obtain $\Omega(n) = o(n \log n \log\log n)^\kappa$.

Proof. We first observe that (γ) implies

$$3.2(1) \qquad 1 \leqslant \Delta\lambda_n/\Delta\lambda_{n+1} \leqslant M.$$

As in Theorem 2(b) we see that the condition 2.3(2) is necessary, further that to prove it sufficient it is enough to derive from it that the sums S_1 and S_2 in 2.3(6) converge to 0 as $\omega \to \infty$. We shall first deduce $S_2 \to 0$ from $S_1 \to 0$. Using the inequality $\Delta(a_n b_n) \geqslant b_{n+1}\Delta a_n$ if $a_n \geqslant 0$, b_n increases, we see that with the λ''_n of 2.3(6),

$$\begin{aligned}\Delta[(\lambda(\omega) - \lambda_{n+2})^{\kappa-1}\Omega(n)] &\geqslant \Omega(n+1)\,\Delta(\lambda(\omega) - \lambda_{n+2})^{\kappa-1}\\ &\geqslant (1-\kappa)\Omega(n)\,(\lambda(\omega) - \lambda_{n+2})^{\kappa-2}\Delta\lambda_{n+2}\\ &\geqslant C\Omega(n)(\lambda(\omega) - \lambda''_n)^{\kappa-2}\Delta\lambda_n.\end{aligned}$$

Therefore, using the formula of partial summation, 2.3(2) and 3.2(1),

$$\begin{aligned}S_2 &\leqslant C_1\lambda(\omega)^{-\kappa}\sum_{n=0}^{\omega_1}\Delta\lambda_{n+2}\Delta[(\lambda(\omega) - \lambda_{n+2})^{\kappa-1}\Omega(n)]\\ &= C_1\lambda(\omega)^{-\kappa}\Big\{-\Delta\lambda_2(\lambda(\omega) - \lambda_2)^{\kappa-1}\Omega(0) + \Delta\lambda_{\omega_1+2}(\lambda(\omega) - \lambda_{\omega_1+2})^{\kappa-1}\Omega(\omega_1+1)\\ &\qquad - \sum_{n=1}^{\omega_1}(\lambda(\omega) - \lambda_{n+2})^{\kappa-1}\Omega(n)\,\Delta^2\lambda_{n+2}\Big\}\\ &\leqslant o(1) + C_1\lambda(\omega)^{-\kappa}\sum_{n=0}^{\omega_1+2}(\lambda(\omega) - \lambda_n)^{\kappa-1}\Omega(n)\left|\Delta^2\lambda_n\right|.\end{aligned}$$

But the second term is $- S_1$ with ω_1 replaced by $\omega_1 + 2$. Thus we have only to show that $S_1 \to 0$ if $\omega_1 < \omega - 1$ or that

$$S' = \sum_{n \leqslant \omega - 1} (\Delta\lambda_n)^{-\epsilon} |\Delta^2\lambda_n| (\lambda(\omega) - \lambda_n)^{\epsilon - 1}$$

is bounded. We break up S' into three parts $\sum_1, \sum_2, \sum_3$ according to the inequalities $n \leqslant \log \omega$, $\log \omega < n \leqslant \frac{1}{2}\omega$, $\frac{1}{2}\omega < n \leqslant \omega - 1$. For $\sum_1$ we have $\lambda(\omega) - \lambda_n \geqslant \lambda(\omega) - \lambda(\log \omega) \to +\infty$ by (β), and therefore

$$\begin{aligned}\sum\nolimits_1 &= o(1) \sum_{n \leqslant \log \omega} (\Delta\lambda_n)^{-\epsilon} |\Delta^2\lambda_n| = o(1) \Big|\sum_{n=0}^{\infty} \Delta(\Delta\lambda_n)^{1-\epsilon}\Big| \\ &= o(1)\,(\Delta\lambda_0)^{1-\epsilon} = o(1).\end{aligned}$$

On the other hand, since $\Delta(\lambda(\log n)/\lambda(n)) \leqslant 0$,

3.2(2) $$\frac{\lambda(\log(n+1)) - \lambda(\log n)}{\Delta\lambda_n} \leqslant \frac{\lambda(\log n)}{\lambda(n)} \leqslant 1.$$

Using (α), 3.2(2) and (β) we see that

$$\begin{aligned}\Delta\lambda(\log n) &= \lambda(\log n + 1) - \lambda(\log n) \\ &\leqslant [\lambda(\log 4n) - \lambda(\log(4n-1))] + \ldots + [\lambda(\log(n+1)) - \lambda(\log n)] \\ &\leqslant 3n[\lambda(\log(n+1)) - \lambda(\log n)] \leqslant C_2 n\,\Delta\lambda_n\end{aligned}$$

and therefore

3.2(3) $$\Delta\lambda(\log n)/(n\,\Delta\lambda_n) \leqslant C_2$$

for some constant C_2. We have further

$$\lambda(\omega) - \lambda_n \geqslant (\omega - n)\,\Delta\lambda(\omega) \geqslant (\omega/2)\,\Delta\lambda(\omega)$$

if $0 \leqslant n \leqslant \frac{1}{2}\omega$. Therefore

$$\begin{aligned}\sum\nolimits_2 &\leqslant C_3(\omega\,\Delta\lambda(\omega))^{\epsilon-1} \sum_{\log \omega \leqslant n \leqslant \omega} (\Delta\lambda_n)^{-\epsilon} |\Delta^2\lambda_n| \\ &\leqslant C_4 \left(\frac{\Delta\lambda(\log \omega)}{\omega\,\Delta\lambda(\omega)}\right)^{1-\epsilon} = O(1),\end{aligned}$$

by 3.2(3). Finally,

$$\sum\nolimits_3 \leqslant (\Delta\lambda(\omega))^{\epsilon-1} \sum_{\frac{1}{2}\omega < n \leqslant \omega - 1} (\Delta\lambda_n)^{-\epsilon} |\Delta^2\lambda_n| \leqslant C_5 \left(\frac{\Delta\lambda(\frac{1}{2}\omega)}{\Delta\lambda(\omega)}\right)^{1-\epsilon} = O(1)$$

by (γ). This completes the proof.

4. Absolute summability functions

4.1. Let $\Omega(n)$ be, as before, a non-decreasing positive function which tends to $+\infty$ with n. In analogy with our former definitions we shall say that $\Omega(n)$ is an *absolute summability function of a method of summation* A (given by 1.1(1)), if any bounded sequence s_n for which $s_n = 0$ except for a subsequence $\{n_\nu\}$ with the counting function $\omega(n) \leqslant \Omega(n)$, is absolutely A-summable, that is if $\sum |\sigma_m - \sigma_{m-1}| < +\infty$ for any such sequence.

The following Lemma will be useful. (With another proof, the Lemma has been communicated to the author by Dr. K. Zeller, Tübingen).

LEMMA 1. *The transformation*

$$v_m = \sum_{\nu=0}^{\infty} b_{m\nu} s_\nu \qquad (m = 0, 1, \dots) \tag{4.1(1)}$$

maps any bounded sequence $s = \{s_\nu\}$ into a sequence $v = \{v_m\}$ with $\sum |v_m| < +\infty$ if and only if one of the following three conditions is fulfilled:

$$\Big| \sum_{m \in e_1} \sum_{\nu \in e} b_{m\nu} \Big| \leqslant M, \tag{4.1(2)}$$

$$\sum_{m=0}^{\infty} \Big| \sum_{\nu \in e} b_{m\nu} \Big| \leqslant M, \tag{4.1(3)}$$

$$\sum_{m=0}^{\infty} \Big| \sum_{\nu \in E} b_{m\nu} \Big| \leqslant M. \tag{4.1(4)}$$

Here E is an arbitrary subset and e, e_1 arbitrary finite subsets of the set of all positive integers, and the M independent of e, e_1, E.

Proof. The conditions are equivalent. It is clear, that 4.1(4) implies 4.1(3) and this imples 4.1(2), and we leave to the reader the elementary proof that 4.1(2) implies 4.1(4). Further, $\sum_{\nu=0}^{\infty} |b_{m\nu}| < +\infty$, $m = 0, 1, \dots$ is necessary and is also a consequence of any of our conditions.

Let S and V be Banach spaces of bounded sequences $s = \{s_\nu\}$ and of sequences $v = \{v_m\}$ with $\sum |v_m| < +\infty$, respectively. Suppose that $v = B(s)$, defined by 4.1(1), maps S into V. For a fixed m, $\sum_\nu b_{m\nu} s_\nu$ is a linear functional in S. Therefore the transformation $v = B_m(s)$, defined by $v_\mu = \sum_{\nu=0}^{\infty} b_{\mu\nu} s_\nu$ for $0 \leqslant \mu \leqslant m$, $v_\mu = 0$ for $\mu > m$, is a linear operation mapping S into V. But clearly $B_m(s) \to B(s)$ for $s \in S$ in the norm of the space V. Therefore $v = B(s)$ is also a linear operation and there is an M such that $\|v\| \leqslant M \|s\|$. But this is identical with 4.1(4), if we put $s_\nu = 1$ for $\nu \in E$, $s_\nu = 0$ for $\nu \overline{\in} E$.

It remains to show that if 4.1(4) is true, then $v = B(s)$ maps S into V. The function $F(s) = \sum_{m=0}^{\infty} \left| \sum_{\nu=0}^{\infty} b_{m\nu} s_\nu \right| \leqslant +\infty$ is clearly lower semi-continuous in S. If the sequence $s = \{s_\nu\}$ is positive, takes only a finite number of values and if $\|s\| \leqslant 1$, then $s = a^{(1)} s^{(1)} + \dots + a^{(p)} s^{(p)}$, where the $s^{(i)}$ are sequences of 0's and 1's, and $a^{(i)} \geqslant 0$, $\sum a^{(i)} \leqslant 1$. Using 4.1(4) we obtain $F(s) \leqslant \sum a^{(i)} F(s^i) \leqslant M$. Without the condition of positiveness of s we have $F(s) \leqslant 2M$. But these new s are dense in the unit sphere of S. Therefore $F(s) \leqslant 2M$ for any s with $\|s\| \leqslant 1$, and $F(s) < +\infty$ everywhere. This completes the proof of the Lemma.

4.2. From Lemma 1 we obtain

THEOREM 6. *In order that $\Omega(n)$ be an absolute summability function of the method* 1.1(1) *for which $\sum |a_{0n}| < +\infty$, it is necessary and sufficient that for for any finite or infinite sequence $n_1 < n_2 < \dots$ with the counting function $\omega(n) \leqslant \Omega(n)$ there is an M such that*

4.2(1) $$\operatorname*{var}_{m} \sum_{\nu=1}^{\infty} a_{mp_\nu} \leqslant M$$

for any subsequence p_ν of the sequence n_ν.

Proof. We apply Lemma 1 to the transformation 4.1(1), where $b_{m\nu}$ is $a_{mn_\nu} - a_{m-1,\,n_\nu}$ and $a_{-1,\,n} = 0$. Then 4.2(1) is equivalent to 4.1(4).

There are of course two other forms of the condition which are obtained from 4.1(2) or 4.1(3). More useful is the following *sufficient* condition:

4.2(2) $$\sum_{\nu=1}^{\infty} \operatorname*{var}_{m} a_{mn_\nu} < +\infty$$

for any sequence $n_1 < n_2 < \ldots$ whose counting function does not exceed $\Omega(n)$.

THEOREM 7. *The method of summation* A *generated by the matrix (a_{mn}) for which $\sum|a_{0n}| < +\infty$ has absolute summability functions if and only if the variation of the n-th column $V_n = \operatorname*{var}_{m} a_{mn}$ converges to 0 for $m \to \infty$.*

Proof. (a) *The condition is sufficient.* Suppose that $V_n \to 0$ for $n \to \infty$. Put $W_n = \max_{p \leqslant n} V_p$, take a sequence n_ν such that $\sum W_{n_\nu} < +\infty$ and denote by $\Omega(n)$ the counting function of $\{n_\nu\}$. If n'_ν is an increasing sequence of integers with the counting function $\omega(n) \leqslant \Omega(n)$, then $n'_\nu \geqslant n_\nu$ for all ν [15, 2.1]. But this implies $\sum V_{n'_\nu} < +\infty$. Applying the sufficient condition 4.2(2) we see that the matrix $A' = (a_{mn'_\nu})$ sums absolutely every bounded sequence, and the matrix A every bounded sequence s_n such that $s_n = 0$ if $n \neq n'_\nu (\nu = 1, 2, \ldots)$. Therefore, $\Omega(n)$ is an absolute summability function for A.

(b) *The condition is necessary.* Suppose that V_n does not tend to 0 and that $\Omega(n)$ is an absolute summability function for the method A. We shall show that there is a sequence n_ν with the counting function $\omega(n) \leqslant \Omega(n)$ such that

4.2(3) $$\operatorname*{var}_{m} \sum_{\nu=1} a_{mn_\nu} = +\infty.$$

This contradiction with Theorem 6 will show that no absolute summability function $\Omega(n)$ can exist.

If the integer p is sufficiently large, the sequence consisting of p alone has certainly the counting function $\leqslant \Omega(n)$; therefore 4.2(1) shows that almost all V_n are finite. We write $b_{mn} = a_{mn} - a_{m-1,\,n}$ $(a_{-1,\,n} = 0)$. Then for any sequence n_ν with the counting function $\leqslant \Omega(n)$ all series $\sum_{\nu=1}^{\infty} b_{mn_\nu} s_{n_\nu}$, $m = 0, 1, \ldots$ must converge for all bounded s_{n_ν}. It follows that all series $\sum_\nu |b_{mn_\nu}|$ converge. It is now clear that there is a monotone sequence of integers p_r whose counting function is $\leqslant \Omega(n)$, such that all series $\sum_m |b_{mp_r}|$ and $\sum_r |b_{mp_r}|$ are convergent and that

4.2(4) $$\sum_m |b_{mp_r}| \geqslant \epsilon \qquad (r = 1, 2, \ldots)$$

for some constant $\epsilon > 0$. For simplicity we write $c_{m\nu}$ instead of b_{mp_ν}. Inductively we choose two increasing sequences of integers r_ν, M_ν. If all numbers with indices less than ν are defined, we choose first an $M_\nu > M_{\nu-1}$ which satisfies

4.2(5) $$A_\nu = \sum_{m > M_\nu} \sum_{\mu=1}^{\nu-1} |c_{mr_\mu}| < \epsilon/5,$$

then $r_\nu > r_{\nu-1}$ such that

4.2(6) $$B_\nu = \sum_{m \leq M_\nu} \sum_{r \geq r_\nu} |c_{mr}| < \epsilon/5.$$

We have then

$$\sum_{M_\nu < m \leq M_{\nu+1}} \Big| \sum_{\mu=1}^{\infty} c_{mr_\mu} \Big| \geq \sum_{M_\nu < m \leq M_{\nu+1}} |c_{mr_\nu}| - A_\nu - B_{\nu+1}$$

$$\geq \sum_{m=0}^{\infty} |c_{mr_\nu}| - \sum_{m \leq M_\nu} |c_{mr_\nu}| - \sum_{m > M_{\nu+1}} |c_{mr_\nu}| - 2\epsilon/5$$

$$\geq \epsilon - 4\epsilon/5 = \epsilon/5$$

by 4.2(5), 4.2(6), and 4.2(4). It follows that $\sum_m \left|\sum_\nu c_{nr_\nu}\right| = +\infty$, and this proves 4.2(3). The proof is complete.

4.3. As an example of application of Theorem 7 we consider Abel, Riesz and Hausdorff methods.

(i) The method $A(\lambda_n)$ has absolute summability functions if it has summability functions, that is if and only if $\Delta\lambda_n/\lambda_n \to 0$ (compare §2.1).

In fact, the coefficient $a_n(x) = e^{-\lambda_n x} - e^{-\lambda_{n+1} x}$ of the $A(\lambda_n)$ transformation 2.1(4) has its maximum for some value x_n of x between λ_n^{-1} and λ_{n+1}^{-1}, and is monotone in $0 \leq x \leq x_n$ and $x \geq x_n$. Therefore,

$$V_n = \operatorname*{var}_{0 \leq x < +\infty} a_n(x) = 2a_n(x_n) \to 0, \qquad n \to \infty,$$

if $A(\lambda_n)$ has summability functions of the first kind. This proof applies also to $R(\lambda_n, \kappa)$, $\kappa > 0$ and gives the same result (in fact, to any regular method A for which a_{mn} has one single maximum in every column).

(ii) A regular Hausdorff method H_g with the generating function $g(t)$ of bounded variation has absolute summability functions whenever H_g has summability functions, that is if and only if $g(t)$ is continuous at $t = 1$ [14, Theorem 13].

For the method H_g,

$$a_{mn} = \int_0^1 p_{nm}(t)dg(t), \quad p_{nm}(t) = \binom{m}{n} t^n (1-t)^{m-n}, \qquad 0 \leq n \leq m,$$

and $a_{mn} = 0$ for $n > m$. Therefore, if H_g has summability functions,

4.3(1) $$V_n = \operatorname*{var}_m a_{mn} \leq |a_{nn}| + \int_0^1 \sum_{m=n}^{\infty} |p_{nm}(t) - p_{n,\,m+1}(t)|\, |dg(t)|$$

$$= o(1) + \int_0^1 P(t) |dg(t)|,$$

say. But for fixed n and t, $p_{nm}(t)$ is first increasing with m and then decreasing, the maximal value being $O(n^{-\frac{1}{2}}) = o(1)$ for $n \to \infty$ uniformly in any interval $\delta \leqslant t \leqslant 1 - \delta$, $\delta > 0$. Moreover $P_n(t) \leqslant 2$ for all n and t. Since $g(t)$ is continuous at $t = 0$ (by the regularity of H_g) and at $t = 1$, 4.3(1) implies $V_n \to 0$, which proves our result.

4.4. In this and the next section we use conditions 4.2(1) and 4.2(2) to find all absolute summability functions of the Cesàro, Euler-Knopp and Borel methods.

THEOREM 8. *A function $\Omega(n)$ is an absolute summability function of the method C_α if and only if*

$$\sum_{n=1}^{\infty} n^{-1-\alpha}\,\Omega(n) < +\infty, \qquad 0 < \alpha < 1, \tag{4.4(1)}$$

or

$$\sum_{n=1}^{\infty} n^{-2}\,\Omega(n) < +\infty, \qquad \alpha \geqslant 1. \tag{4.4(2)}$$

We shall need two lemmas.

LEMMA 2. *For a sequence of integers $0 < n_1 < n_2 < \dots$ with the counting function $\omega(n)$ the two following conditions are equivalent* ($\alpha > 0$):

$$\sum_{n=1}^{\infty} n^{-1-\alpha}\,\omega(n) < +\infty \tag{4.4(3)}$$

$$\sum_{\nu=1}^{\infty} n_\nu^{-\alpha} < +\infty. \tag{4.4(4)}$$

In fact,

$$\begin{aligned}\sum n^{-1-\alpha}\,\omega(n) &= \sum_{n=1}^{\infty}\sum_{n_\nu \leqslant n} n^{-1-\alpha} = \sum_{\nu=1}^{\infty}\sum_{n \geqq n_\nu} n^{-1-\alpha}\\ &= \theta \sum_{\nu=1}^{\infty} n_\nu^{-\alpha},\end{aligned}$$

where θ is some number, contained in a fixed interval (a, b), $0 < a < b < \infty$.

LEMMA 3. *Let $\sum n^{-1-\alpha}\Omega(n) = +\infty$, $\alpha > 0$ and let $a > 1$ be an integer. Set $p_\nu = a^\nu$. Then*

$$\sum_{\nu=1}^{\infty} p_\nu^{-\alpha}[\Omega(p_\nu) - \Omega(p_{\nu-1})] = +\infty. \tag{4.4(5)}$$

For we have, with positive constants C_1, C_2,

$$\begin{aligned}&\sum_{\nu=1}^{N} p_\nu^{-\alpha}[\Omega(p_\nu) - \Omega(p_{\nu-1})]\\ &\qquad = -\,\Omega(p_0)p_1^{-\alpha} + \sum_{\nu=1}^{N-1}\Omega(p_\nu)\,(p_\nu^{-\alpha} - p_{\nu+1}^{-\alpha}) + p_N^{-\alpha}\Omega(p_N)\end{aligned}$$

$$\geqslant O(1) + C_1 \sum_{\nu=1}^{N} \Omega(p_\nu) p_\nu^{-\alpha}$$

$$\geqslant O(1) + C_2 \sum_{\nu=1}^{N-1} \Omega(p_\nu) \sum_{n=p_{\nu-1}}^{p_\nu - 1} n^{-1-\alpha}$$

$$\geqslant O(1) + C_2 \sum_{n=1}^{p_{N-1}-1} n^{-1-\alpha}\Omega(n).$$

Proof of Theorem 8. (a) *The conditions are sufficient.* Suppose that 4.4(1) holds with some α, $0 < \alpha \leqslant 1$, and let $n_1 < n_2 < \dots$ have a counting function $\omega(n) \leqslant \Omega(n)$. Then $\sum n_\nu^{-\alpha} < +\infty$, by Lemma 2. It will be sufficient to show that 4.2(2) holds. But for the method C_α, $a_{mn} = 0$ for $m < n$,

4.4(6) $$a_{mn} = (A_m^\alpha)^{-1} A_{m-n}^{\alpha-1} \text{ for } m \geqslant n,\ A_n^\alpha = \binom{n+\alpha}{n} \cong n^\alpha/\Gamma(\alpha+1),$$

and a_{mn} is a decreasing function of m for $m \geqslant n$. Therefore,

$$\operatorname*{var}_m a_{mn} = 2a_{nn} = 2(A_n^\alpha)^{-1} \leqslant Cn^{-\alpha},$$

and 4.2(2) follows. The rest follows from the inclusion $|C_\alpha| \subset |C_\beta|$ for $\alpha \leqslant \beta$.

(b) *The conditions are necessary.* First suppose $0 < \alpha < 1$. By [15, 5.1] we may assume that $\Omega(n) = o(n)$. Suppose that $\sum n^{-1-\alpha}\Omega(n) = +\infty$. We define $\omega_1(n)$ inductively by putting $\omega_1(1) = 0$ and, if $\omega_1(n)$ is known, $\omega_1(n+1) = \omega_1(n) + 1$ if this number is $\leqslant \Omega(n+1)$, and $\omega_1(n+1) = \omega_1(n)$ in the contrary case. Using $\Omega(n) = o(n)$ one proves easily that $\sum n^{-1-\alpha}\omega_1(n) = +\infty$. $\omega_1(n)$ is the counting function of some sequence. Omitting, if necessary, some terms of this sequence, we obtain another sequence of integers $n_1 < n_2 < \dots$ such that (i) its counting function $\omega(n) \leqslant \Omega(n)$; (ii) $\sum n_\nu^{-\alpha} = +\infty$; (iii) for any ν, $n_\nu + 1$ does not belong to the sequence. We now observe that the coefficient a_{mn} given by 4.4(6) is decreasing for $m \geqslant n$ and that

$$a_{nn} - a_{n+1,n} = (A_n^\alpha)^{-1} - (A_{n+1}^\alpha)^{-1} A_1^{\alpha-1}$$

$$= (A_n^\alpha)^{-1} \frac{(1-\alpha)n + 1}{n+1+\alpha} \geqslant Cn^{-\alpha}$$

with some constant $C > 0$. Using (iii) and (ii) we obtain

$$\operatorname{var} \sum_{\nu=1}^{\infty} a_{mn_\nu} \geqslant \sum_{\mu=1}^{\infty} \sum_{\nu=1}^{\infty} (a_{n_\mu n_\nu} - a_{n_\mu+1,\, n_\nu})$$

$$\geqslant \sum_{\mu=1}^{\infty} (a_{n_\mu n_\mu} - a_{n_\mu+1,\, n_\mu}) \geqslant C\sum n_\mu^{-\alpha} = +\infty,$$

and the result follows by Theorem 6.

Next consider the case $\alpha \geqslant 1$. We may assume $\alpha > 1$. Without restriction of generality we may also suppose that $\Omega(n) = o(n)$ and takes only integral values. We choose $k > e\alpha$ and then an integer $a > k\alpha$. If 4.4(2) is not ful-

filled, we must have $\sum_{\nu=1}^{\infty} p_\nu^{-1} q_\nu = +\infty$, $q_\nu = \Omega(p_\nu) - \Omega(p_{\nu-1})$, by Lemma 3. Consider the sequence consisting of all groups of integers n, $p_\nu \leqslant n < p_\nu + q_\nu$ $(\nu = 1, 2, \ldots)$. The counting function of the sequence is $\leqslant \Omega(n)$. Put $f(m) = \sum_{\nu=1}^{\infty} f_\nu(m)$, $f_\nu(m) = \sum_{p_\nu \leqslant n < p_\nu + q_\nu} a_{mn}$. If we can show that

4.4(7) $$\operatorname*{var}_m f(m) = +\infty$$

our result will follow by Theorem 6. Since

$$a_{m+1,n}\, a_{mn}^{-1} - 1 = \frac{an - m - 1}{(m - n + 1)(m + a + 1)}, \qquad m \geqslant n,$$

the coefficient a_{mn} is surely decreasing as a function of m for $m > a\,n$. Therefore, $f_\nu(m)$ decreases if $m > a(p_\nu + q_\nu)$. Let $m'_\nu = [ap_\nu]$, $m''_\nu = [kap_\nu]$. Since $m''_\nu < p_{\nu+1}$, $f_\mu(m) = 0$ for $\mu > \nu$, $m \leqslant m''_\nu$. On the other hand, $f_\mu(m)$, $\mu < \nu$ are decreasing for $m \geqslant m'_\nu$. Therefore

4.4(8) $$f(m'_\nu) - f(m''_\nu) \geqslant f_\nu(m'_\nu) - f_\nu(m''_\nu).$$

Using 4.4(6) and $q_\nu = o(p_\nu)$ we have

4.4(9) $$\begin{aligned} f_\nu(m'_\nu) &= \sum_{p_\nu \leqslant n < p_\nu + q_\nu} a_{m'_\nu, n} \geqslant q_\nu a_{m'_\nu,\, p_\nu + q_\nu} \\ &\cong Cq_\nu (ap_\nu)^{-a}((a-1)p_\nu)^{a-1} \geqslant Cq_\nu(eap_\nu)^{-1}, \end{aligned}$$

where C denotes the constant $\Gamma(a+1)/\Gamma(a)$. On the other hand

4.4(10) $$\begin{aligned} f_\nu(m''_\nu) &\leqslant q_\nu a_{m''_\nu p_\nu} \cong Cq_\nu(kap_\nu)^{-a}((ka-1)p_\nu)^{a-1} \\ &\leqslant Cq_\nu(kp_\nu)^{-1}. \end{aligned}$$

Since $k > ea$, from 4.4(8), 4.4(9), and 4.4(10) it follows that

$$f(m'_\nu) - f(m''_\nu) \geqslant C_1 q_\nu p_\nu^{-1}, \qquad C_1 > 0,$$

and we obtain 4.4(7).

We do not know whether the condition 4.4(2), which is clearly necessary, is also sufficient for the Abel method A. But there is a proof similar to the last case of Theorem 8 if $q_\nu p_\nu^{-1}$ is sufficiently smooth, if for instance $\Omega(n)$ is a quotient of n by iterated logarithms.

4.5. THEOREM 9. *A function* $\Omega(n)$ *is an absolute summability function of the Euler-Knopp method* E_t, $0 < t < 1$, *or of the Borel method* B *if and only if*

4.5(1) $$\sum_{n=1}^{\infty} n^{-3/2}\Omega(n) < +\infty.$$

Proof. In view of the inclusion $|\mathrm{E}_t| < |\mathrm{B}|$ (Knopp-Lorentz [11]) it will be sufficient to show that (i) 4.5(1) is sufficient for the method E_t; (ii) 4.5(1) is necessary for B.

Now the E_t transformation is

$$\sigma_m = \sum_{n=0}^{m} p_{nm}(t)s_n \qquad (m = 0, 1, \ldots).$$

For fixed n and t, $p_{nm}(t)$ takes its maximal value at $m = m_0$, where m_0 is the least integer satisfying $m > nt^{-1} - 1$. This maximum is $\leqslant C(t)n^{-\frac{1}{2}}$. As $p_{nm}(t)$ is monotone in $n \leqslant m \leqslant m_0$ and $m \geqslant m_0$,

4.5(2) $$\operatorname*{var}_m p_{nm}(t) \leqslant 2C(t)n^{-\frac{1}{2}}.$$

Now if $\{n_\nu\}$ is a sequence with the counting function $\omega(n) \leqslant \Omega(n)$, we have $\sum n_\nu^{-\frac{1}{2}} < +\infty$ by Lemma 2, and from 4.5(2) we see that 4.2(2) holds. This proves (i).

Now suppose the series 4.5(1) be divergent. Taking $a = 4$ we apply Lemma 3 and obtain $\sum p_\nu^{-\frac{1}{2}}q_\nu = +\infty$ with $q_\nu = \Omega(p_\nu) - \Omega(p_{\nu-1})$. Again we may assume that $\Omega(n)$ takes only integral values and [15, 5.2] has the property $\Omega(n) = o(n^{\frac{1}{2}})$. Consider the sequence (with counting function $\leqslant \Omega(n)$) which consists of all integers n contained in the intervals $p_\nu \leqslant n < p_\nu + q_\nu$, $(\nu = 1, 2, \ldots)$. Let

$$f(x) = \sum_{\nu=1}^{\infty} f_\nu(x), \quad f_\nu(x) = \sum_{p_\nu \leqslant n < p_\nu + q_\nu} e^{-x}x^n/n!$$

To prove (ii) we have, by Theorem 6 (or rather its continuous analogue), to show that

4.5(3) $$\operatorname*{var}_{0 \leqslant x < +\infty} f(x) = +\infty.$$

But $a_n(x) = e^{-x}x^n/n!$ attains its maximum $\cong (2\pi n)^{-\frac{1}{2}}$ at $x = n$. Moreover, if $0 \leqslant r \leqslant Cn^{-\frac{1}{2}}$, then $a_{n+r}(n) \geqslant C_1 n^{-\frac{1}{2}}$. Since $q_\nu = o(p_\nu^{-\frac{1}{2}})$, we obtain

$$f(p_\nu) \geqslant f_\nu(p_\nu) \geqslant C_1 p_\nu^{-\frac{1}{2}} q_\nu.$$

On the other hand,

$$f(3p_\nu) = \sum_{\mu=1}^{\infty} f_\mu(3p_\nu) \leqslant \sum_{|n-3p_\nu| \geq p_\nu} a_n(3p_\nu) = O(e^{-\gamma p_\nu})$$

for some $\gamma > 0$ (see for instance [5, p.200]). We see that

$$\operatorname*{var}_x f(x) \geqslant \sum_{\nu=1}^{\infty} \{f_\nu(p_\nu) + O(e^{-\gamma p_\nu})\} \geqslant C_1 \sum p_\nu^{-\frac{1}{2}} q_\nu + O(1) = +\infty,$$

which proves 4.5(3).

5. Some further theorems, applications and remarks

5.1. In this section we wish to discuss some applications of the results in [14], [15] and this paper and their relation to known theorems. We begin with the following remark. The definition of a summability function of the second

kind (see §1.1) may obviously be restated as follows: $\Omega(n)$ is a summability function of the second kind of a regular A if and only if $\sigma_n = (s_0+s_1+\ldots+s_n)/(n+1) = s + O(n^{-1}\Omega(n))$ implies the A-summability of s_n to s. Thus from [15, 5.2] follows the theorem of Knopp ([10], also [5, p. 213]): $\sigma_n = s + o(n^{-\frac{1}{2}})$ implies E_t -summability of s_n together with the result that this is the best possible theorem.

5.2. We observed in [15, 3.1] that summability functions may be used to show that Tauberian conditions of a certain kind may not be improved. Thus our results in §2 and §3 imply that under certain conditions $u_n = O(\Delta\lambda_n/\lambda_n)$ is the best possible Tauberian condition for $R(\lambda_n, \kappa)$ and $A(\lambda_n)$. This method however fails to give the full truth if $\Delta\lambda_n/\lambda_n$ is smaller than n^{-1}, since a regular method of summation cannot possess summability functions like $n \log n$. The following theorem, based on the sufficiency part of [14, Theorem 8], gives, as far as we know, a precise result for all practically interesting special methods of summation (compare also [12]).

THEOREM 10. (i) *Suppose that* $A = (a_{mn})$ *is a regular method of summation and* $n_1 < n_2 < \ldots$ *a sequence of integers for which*

5.2(1) $$\lim_{m\to\infty} \Big\{\max_{\nu} \sum_{n=n_\nu}^{n_{\nu+1}-1} |a_{mn}|\Big\} = 0.$$

Then $u_n = 0$ *for* $n \neq n_\nu$ *is not a Tauberian condition for* A.

(ii) *If, moreover,* $c_n \to 0$, $c_n \geqslant 0$ *and*

5.2(2) $$\sum_{n=n_\nu}^{n_{\nu+1}-1} c_n \geqslant \delta > 0 \qquad (\nu = 1, 2, \ldots),$$

then $u_n = O(c_n)$ *is not a Tauberian condition for* A.

Both statements are true even for bounded sequences $s_n = \sum_{\nu=0}^{n} u_\nu$.

Proof. Let $A' = (a'_{m\nu})$ and $a'_{m\nu} = \sum_{n=n_\nu}^{n_{\nu+1}-1} a_{mn}$. Then $\max_\nu a'_{m\nu} \to 0$ for $m \to \infty$, and by [14, Theorem 8 and 8*], there is a bounded divergent sequence which is A' -summable. This implies (i).

To prove (ii) consider the method $A'' = (a''_{m\nu})$, where

5.2(3) $$a''_{m\nu} = \sum_{n=n_\nu}^{n_{\nu+1}-1} |a_{mn}|.$$

Since $\max_\nu a''_{m\nu} \to 0$ and $\sum_\nu |a''_{m\nu}| < +\infty$ for any m, by [14, Theorem 8], there is a divergent sequence of 0's and 1's A'' -summable to 0 (Theorem 8 is formulated for regular methods, but only the two properties of A'' stated above are used in the proof). In other words there is a subsequence $\nu(k)$ of the ν such that

5.2(4)
$$\sum_{k=1}^{\infty} \sum_{n=n_{\nu(k)}}^{n_{\nu(k)+1}-1} |a_{mn}| \to 0 \qquad \text{for } m \to \infty.$$

Using 5.2(2) and $c_n \to 0$ we can choose, for all large k, an n'_k between $n_{\nu(k)}$ and $n_{\nu(k)+1}$ and u_n positive in $n_{\nu(k)} \leqslant n < n'_k$, negative in $n'_k \leqslant n < n_{\nu(k)+1}$ such that

$$u_n = O(c_n), \quad \sum_{n_{\nu(k)}}^{n'_k-1} u_n = \delta/3, \quad \sum_{n_{\nu(k)}}^{n_{\nu(k)+1}-1} u_n = 0.$$

We put $u_n = 0$ for the remaining n. The sequence $s_n = \sum_{p=0}^{n} u_p$ is bounded, divergent, A-summable to 0 and has the property $u_n = O(c_n)$. This proves (ii).

It follows from the proof that Theorem 10 remains true if instead of 5.2(1) we assume only

5.2(5)
$$\lim_{m \to \infty} \{\max_{r} a''_{m\nu_r}\} = 0, \qquad m \to \infty$$

for a subsequence ν_r of the ν.

5.3. From the possible applications of Theorem 10 we choose those to Riesz and Wiener methods.

THEOREM 11. *Suppose that $\lambda(n) = \lambda$ is a positive function increasing to $+\infty$ with n.*

(i) *If n_ν is a sequence of integers increasing to $+\infty$ and such that $\lim_{\nu\to\infty} [\lambda(n_{\nu+1})/\lambda(n_\nu)] = 1$, then $u_n = 0$, $n \neq n_\nu$ is not a Tauberian condition of the method* $R(\lambda_n, \kappa)$, $\kappa > 0$.

(ii) *If $c_n = \varphi(n)\Delta\lambda_n/\lambda_n \to 0$, where $\sum c_n = +\infty$ and $\varphi(n) \to +\infty$, then $u_n = O(c_n)$ is not a Tauberian condition for* $R(\lambda_n, \kappa)$.

Proof. We may assume $0 < \kappa < 1$. By 2.1(3) we have

5.3(1)
$$a''_\nu(\omega) = \sum_{n_\nu \leqslant n < n_{\nu+1}} a_n(\omega)$$
$$= \begin{cases} \lambda(\omega)^{-\kappa}\{[\lambda(\omega)-\lambda(n_\nu)]^\kappa - [\lambda(\omega)-\lambda(n_{\nu+1})]^\kappa\} & \text{if } \omega \geqslant n_{\nu+1} \\ \lambda(\omega)^{-\kappa}[\lambda(\omega)-\lambda(n_\nu)]^\kappa & \text{if } n_\nu \leqslant \omega < n_{\nu+1}, \\ 0 & \text{if } \omega < n_\nu. \end{cases}$$

Using the inequality $0 < \kappa < 1$ we see that for fixed ν, $a''_\nu(\omega)$ takes its maximum for $\omega = n_{\nu+1}$ which is equal to $\lambda(n_{\nu+1})^{-\kappa}[\lambda(n_{\nu+1})-\lambda(n_\nu)]^\kappa$. Since the lower limit of this expression for $\nu \to \infty$ is 0, and since $a''_\nu(\omega) \to 0$ for fixed ν and $\omega \to \infty$, there is a subsequence ν_r such that 5.2(5) holds. Using the remark at the end of 5.2 we obtain (i).

In proving (ii) we may suppose that $c_n \leqslant 1$. We take n_1 arbitrary and define $n_{\nu+1}$, if n_ν is known, to be the first integer $> n_\nu$ such that $\sum_{n_\nu \leqslant n < n_{\nu+1}} c_n \geqslant 1$. Then

$$2 \geqslant \sum_{n_\nu \leqslant n < n_{\nu+1}} c_n \geqslant \varphi(n_\nu)\lambda(n_{\nu+1})^{-1}[\lambda(n_{\nu+1}) - \lambda(n_\nu)],$$

and therefore $\lambda(n_{\nu+1})/\lambda(n_\nu) \to 1$. As in the proof of (i) we see that this implies 5.2(1). The proof is completed by applying Theorem 10.

By a different and more difficult method, Theorem 11, (ii) had been proved by Ingham [8]. Instead of our hypothesis $c_n \to 0$ Ingham assumes that $\lambda_{n+1}/\lambda_n \to 1$. This difference is inessential, as in the latter case we may always replace $\varphi(n)$ by a smaller function tending to $+\infty$, for which $c_n \to 0$ holds.

Passing to Wiener's methods, we call a bounded function $f(x)$, $0 \leqslant x < +\infty$ summable to s by a Wiener method W_g, if $\int_0^{+\infty}|g(t)|dt < +\infty$ and

5.3(2) $$\frac{1}{x}\int_0^\infty g\left(\frac{t}{x}\right)f(t)dt \to s\int_0^\infty g(t)dt, \qquad x \to \infty.$$

The well known Tauberian theorem of Pitt [5, p. 296, Theorem 233] asserts that if $\int_0^\infty g(t)t^{ix}dt \neq 0$ for real x, then

5.3(3) $$f(x+\delta) - f(x) \to 0 \qquad \text{for } \delta > 0,\ \delta/x \to 0,\ x \to \infty$$

is a Tauberian condition for the method Wg. In particular, if $f(x)$ is absolutely continuous,

5.3(4) $$f'(x) = O(x^{-1}), \qquad x \to \infty$$

is a Tauberian condition. We use the analogue of Theorem 10 for integrals to show that these conditions cannot be improved.

THEOREM 12. *The conditions*

5.3(5) $$f(x+\delta) - f(x) \to 0 \qquad \text{for } \delta > 0,\ \delta\varphi(x)/x \to 0,\ x \to \infty$$

or

5.3(6) $$f'(x) = O(x^{-1}\varphi(x)), \qquad x \to \infty$$

where $\varphi(x)$ is bounded in any finite interval and $\varphi(x) \to \infty$ are not Tauberian conditions for any method W_g.

Proof. It will be sufficient to consider 5.3(6). We define t_ν, $(\nu = 1, 2, \ldots)$ inductively by $t_1 = 1$,

5.3(7) $$\int_{t_\nu}^{t_{\nu+1}} x^{-1}\varphi(x)dx = 1 \qquad (\nu = 1, 2, \ldots).$$

Then $t_{\nu+1}/t_\nu \to 1$. The expression corresponding to 5.2(3) is

$$a''_\nu(x) = \frac{1}{x}\int_{t_\nu}^{t_{\nu+1}}|g(t/x)|dt = \int_{x^{-1}t_\nu}^{x^{-1}t_{\nu+1}}|g(u)|du.$$

Taking $A > 0$ so large that $\int_A^\infty|g|du < \epsilon$, we observe that the maximal length of $(x^{-1}t_\nu, x^{-1}t_{\nu+1})$ for all ν with $x^{-1}t_\nu \leqslant A$ tends to 0 as $x \to \infty$. This implies that $a''_\nu(x) < \epsilon$ for all ν and all sufficiently large x. Thus we obtain 5.2(1) and 5.3(7) gives the condition 5.2(2) of Theorem 10. The proof is complete.

A theorem on absolute summability corresponding to Theorem 10, (i) may be obtained using Theorem 7, §4.2 instead of [14, Theorem 8]. In this way we obtain that $u_n = 0$, $n \neq n_\nu$ $(\nu = 1, 2, \ldots)$ is not a Tauberian condition for absolute summability by the matrix $A = (a_{mn})$ if

5.3(8) $$\lim_{\nu\to\infty} \{\operatorname{var}_m \sum_{n_\nu \leq n < n_{\nu+1}} a_{mn}\} = 0.$$

(More precisely, if 5.3(8) holds, there are bounded divergent sequences with $u_n = 0$, $n \neq n_\nu$, which are absolutely A -summable.) As an example we have that the high indices theorem for absolute Abel summability of Zygmund [17] cannot be improved.

5.4. In [15, 6.2] it has been shown that $u_n = o(n^{-1})$ is a Tauberian condition for any regular Hausdorff method H_g. We show now that for an unspecified generating function $g(t)$ this condition cannot be improved. *There are regular methods* H_g *such that* $u_n = O(n^{-1})$ *is not a Tauberian condition, even for bounded sequences.*

Set
$$g(t) = \begin{cases} 0 & \text{in } [0, \tfrac{1}{3}), \\ \tfrac{1}{2} & \text{in } [\tfrac{1}{3}, \tfrac{2}{3}), \\ 1 & \text{in } [\tfrac{2}{3}, 1]. \end{cases}$$

The corresponding H_g transformation is given by

5.4(1) $$\sigma_n = \tfrac{1}{2} \sum_{\nu=0}^{n} \binom{n}{\nu} [t_1^\nu (1 - t_1)^{n-\nu} + t_2^\nu (1-t_2)^{n-\nu}] s_\nu, \qquad t_1 = \tfrac{1}{3}, t_2 = \tfrac{2}{3}.$$

Using the well known properties of the Newton probabilities $p_{n\nu}(t) = \binom{n}{\nu} t^\nu (1 - t)^{n-\nu}$ it is easy to prove that under the hypotheses $u_n = O(n^{-1})$, $s_n = O(1)$ the method 5.4(1) is equivalent to the method defined by

5.4(2) $$\sigma_n = \tfrac{1}{2}(s_{[n/3]} + s_{[2n/3]}).$$

Therefore it is sufficient to give a function $s(u)$ of the real argument $u \geqslant 1$ such that $s(u) = O(1)$, $s(u + 1) - s(u) = O(u^{-1})$ and $s(u) + s(2u) \to 0$. But a function of this kind is defined by

$$s(u) = \begin{cases} (-1)^\nu(\log_2 u - \nu) & \text{for } 2^\nu \leqslant u < 2^{\nu+\frac{1}{2}} \\ (-1)^\nu(\nu+1-\log_2 u) & \text{for } 2^{\nu+\frac{1}{2}} \leqslant u < 2^{\nu+1}, \end{cases} \qquad (\nu = 0, 1, \ldots).$$

Our proof in [15, 6.2] was based on a gap theorem of Agnew [2] for the methods H_g. It is perhaps worth while to remark that the following improvement of Agnew's result is true. *For any regular method* H_g *there is a constant* $\lambda = \lambda_g > 1$ *such that* $u_n = 0$ *for* $n \neq n_\nu (\nu = 1, 2, \ldots)$ *is a Tauberian condition for the method* H_g, *if*

5.4(3) $$n_{\nu+1}/n_\nu \geqslant \lambda.$$

(Agnew assumes $n_{\nu+1}/n_\nu \to \infty$ instead of this.) The proof is obtained by combining Agnew's argument with a well known elementary Mercerian

theorem ([1], also [16]). It is not known whether we may take λ_ϱ as near to 1 as we please.

5.5. In this section we make some minor remarks, and corrections to earlier papers.

We first observe, that almost convergence [14, 1] may be defined for sequences of elements x_n of a Banach space. We call x_n almost convergent to x, if

$$\left\| x - \frac{x_{n+1} + \ldots + x_{n+\nu}}{\nu} \right\| \to 0 \text{ for } \nu \to \infty \text{ uniformly in } n. \tag{5.5(1)}$$

(This implies that the $\|x_n\|$ are bounded.) We have, for example, the following theorem. ***Any weakly convergent sequence of elements of a uniformly convex Banach space contains a strongly almost convergent subsequence*** (which is therefore strongly C_α -summable for any $\alpha > 0$). In fact, a modification of the argument used by Kakutani [9] shows that the subsequence x_n which he proves to be strongly C_1-summable, is even strongly almost convergent.

Dr. R. G. Cooke kindly points out that he has used our condition [15, 2.4(1)] for some other purpose in [3]. He also makes the following remark. The condition $\max_n |a_{mn}| \to 0$ is equivalent, for any method A with the property $\sum_n |a_{mn}| \leqslant M$, to the condition

$$\sum_{n=0}^{\infty} a_{mn}^2 \to 0 \qquad m \to \infty, \tag{5.5(2)}$$

for

$$\max_n |a_{mn}|^2 \leqslant \sum_n a_{mn}^2 \leqslant M \max_n |a_{mn}|.$$

Now 5.5(2) is given by Hill [7] as a necessary condition for a method A to possess the Borel property. Hence, by [14, Theorem 8*] if a regular method A has the Borel property, then it possesses summability functions of the first kind.

We note that a theorem by Garabedian, Hille and Wall [4, Theorem 5.2] gives a set of necessary and sufficient conditions in order that all functions $\Omega(n) = o(n)$ be summability functions of the second kind of a Hausdorff method H_ϱ.

We use this opportunity to rectify some mistakes in our previous papers.

In the proof of [14, Theorem 10] the sequence $n_1 < n_2 < \ldots$ depends upon m (it is erroneously stated there that it is the same for all m in question).

In the formulation of Theorem 5 in *Operations in linear metric spaces*, Duke Math. J., vol. 15 (1948) 755-761, replace "when" by "if and only if".

In a review of the above paper (Math. Reviews, vol. 10 (1949), 255) it is stated that the proof of the main Theorem 1 of this paper is incomplete. The slips are, however, of minor nature and are rectified as follows:

(a). The (well known) definition of openness of a mapping is incorrectly formulated on p. 757, lines 1-3. To obtain a correct one, replace the first part of line 3 by: "for any $y \in U_\sigma(y_0)$ an element $x \in U_\epsilon(x_0)$ exists for which $y = Sx$". Only the correct definition is used in the proof.

(b). Lines 15-16 on p. 757 are not sufficient to insure that the set $B_{a,b} = [a < \Phi(y) < b]$ is analytical. But the argument in the text applies to the set $B_b = [\Phi(y) < b]$, and since the $B_{a,b}$ are unions of differences of the B_b, they, too, are analytical.

References

[1] R. P. Agnew, *On equivalence of methods of evaluation of sequences*, Tôhoku Math. J., vol. 35 (1932), 244-252.

[2] ——— *Analytic extension of Hausdorff methods*, Trans. Amer. Math. Soc., vol. 52 (1942), 217-237.

[3] R. G. Cooke, *On mutual consistency and regular T-limits*, Proc. London Math. Soc., vol. (2), 41 (1936), 113-125.

[4] H. L. Garabedian, E. Hille and H. S. Wall, *Formulations of the Hausdorff inclusion problem*, Duke Math. J., vol. 8 (1941), 193-213.

[5] G. H. Hardy, *Divergent series*, Oxford Univ. Press (1949).

[6] G. H. Hardy and M. Riesz, *The general theory of Dirichlet's series*, Cambridge (1915).

[7] J. D. Hill, *Summability of sequences of 0's and 1's*, Ann. of Math., vol. (2) 46 (1945), 556-562.

[8] A. E. Ingham, *Note on the converse of Abel's theorem*, Proc. London Math. Soc., vol. (2) 23 (1924), 326-336.

[9] Shizo Kakutani, *Weak convergence in uniformly convex spaces*, Tôhoku Math. J., vol. 45 (1938), 188-193.

[10] K. Knopp, *Über das Eulersche Summierungsverfahren*, (II), Math. Zeitschr., vol. 18 (1923), 125-156.

[11] K. Knopp and G. G. Lorentz, *Beiträge zur absoluten Limitierung*, Archiv der Math., vol. 2 (1949), 10-16.

[12] G. G. Lorentz, *Beziehungen zwischen den Umkehrsätzen der Limitierungstheorie*, Bericht der Math. Tagung Tübingen, (1946), 97-99.

[13] ——— *Über Limiterungsverfahren, die von einem Stieltjes-Integral abhängen*, Acta Math., vol. 79 (1947), 255-272.

[14] ——— *A contribution to the theory of divergent sequences*, Acta Math., vol. 80 (1948), 167-190.

[15] ———*Direct theorems on methods of summability*, Can. J. of Math., vol. 1 (1949), 305-319.

[16] R. Rado, *Some elementary Tauberian theorems* (I), Quart. J. Math. (Oxford ser.), vol. 9 (1938), 274-282.

[17] A. Zygmund, *On certain integrals*, Trans. Amer. Math. Soc., vol. 55 (1944), 170-204.

The University of Toronto

ctly
part
x''.

=
set
ney,

l. 35
42),
Soc.,
sion
5).
945),
. (2)
l. 45
l. 18
ath.,
richt
ath.,
948),
319.
ol. 9